

Georgia State University
ScholarWorks @ Georgia State University

Psychology Theses

Department of Psychology

8-12-2014

Test Performance: the Influence of Cognitive Load on Reading Comprehension

Amanda Clevinger

Follow this and additional works at: https://scholarworks.gsu.edu/psych_theses

Recommended Citation

Clevinger, Amanda, "Test Performance: the Influence of Cognitive Load on Reading Comprehension." Thesis, Georgia State University, 2014.
https://scholarworks.gsu.edu/psych_theses/123

This Thesis is brought to you for free and open access by the Department of Psychology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Psychology Theses by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.

TEST PERFORMANCE: THE INFLUENCE OF COGNITIVE LOAD ON READING COMPREHENSION

by

AMANDA M. CLEVINGER

Under the Direction of Heather M. Kleider-Offutt

ABSTRACT

Scores from high stakes tests such as the Scholastic Aptitude Test (SAT) are commonly used as criteria for college admission decisions. So, it is of applied importance to identify factors that contribute to susceptibility to failure on these tests. One potential factor addressed in the current study was whether emotional cognitive load differentially impacts those with low working memory capacity or trait anxiety. Individual differences in subjective arousal were also tested as a mechanism contributing to this effect. In Experiment 1, a reading comprehension task revealed that type of cognitive load affected accuracy. In Experiment 2, state anxiety was induced using methods from previous research. The results revealed that, again, only type of cognitive load affected comprehension accuracy. Together, results suggest that arousal induced via disturbing words negatively influence reading performance regardless of superior working memory capacity. These findings are not based on cognitive load in general, but the semantic value of the words processed, in particular, that led to comprehension difficulty. Results are discussed in terms of theoretical and practical implications.

INDEX WORDS: Working memory capacity, Arousal, Trait anxiety, Reading comprehension,
Test performance

TEST PERFORMANCE: THE INFLUENCE OF COGNITIVE LOAD ON READING
COMPREHENSION

by

AMANDA M. CLEVINGER

A Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of

Master of Arts

in the College of Arts and Sciences

Georgia State University

2014

Copyright by
Amanda Marie Clevinger
2014

TEST PERFORMANCE: THE INFLUENCE OF COGNITIVE LOAD ON READING
COMPREHENSION

by

AMANDA M. CLEVINGER

Committee Chair: Heather Kleider-Offutt

Committee: David Washburn
Chris Conway

Electronic Version Approved:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
August 2014

TABLE OF CONTENTS

LIST OF TABLES	vii
LIST OF FIGURES	viii
1 INTRODUCTION.....	1
1.1 Working Memory	1
<i>1.1.1 Working Memory Load.....</i>	<i>1</i>
1.2 Arousal	2
1.3 Current Study	3
2 BACKGROUND.....	5
2.1 Working Memory Capacity.....	5
<i>2.1.1 Baddeley's Model.....</i>	<i>5</i>
<i>2.1.3 Hypotheses Regarding Working Memory</i>	<i>8</i>
2.2 Cognitive Load Theory	9
<i>2.2.1 Cognitive Load and Attention.....</i>	<i>10</i>
<i>2.2.2 Arousal as a Load</i>	<i>10</i>
<i>2.2.3 Hypotheses Regarding Arousal</i>	<i>11</i>
2.3 Trait Anxiety.....	12
<i>2.3.1 Hypotheses Regarding Trait Anxiety</i>	<i>14</i>
2.4 Arousing Stimuli.....	15
<i>2.4.1 Hypotheses Regarding Arousing Stimuli.....</i>	<i>18</i>

2.5	Current Study Overview	19
3	EXPERIMENT 1	21
3.1	Methods.....	21
3.1.1	<i>Participants</i>	21
3.1.2	<i>Materials</i>	21
3.2	Study 1 Results	26
3.2.1	<i>Data Preparation and Variable Calculation</i>	26
3.2.2	<i>Analysis of Variance</i>	28
3.2.3	<i>Additional Analyses</i>	28
3.3	Discussion.....	34
4	EXPERIMENT 2	36
4.1	Methods.....	36
4.1.1	<i>Participants</i>	36
4.1.3	<i>Procedure</i>	37
4.2	Study 2 Results	37
4.2.1	<i>Data Preparation and Variable Calculation</i>	37
4.2.2	<i>Analysis of Variance</i>	39
4.2.3	<i>Additional Analyses</i>	40
4.3	Discussion.....	42
5	GENERAL DISCUSSION.....	45

5.1 Future Directions.....	54
5.2 Conclusion.....	55
REFERENCES.....	57
APPENDICES.....	69
APPENDIX A: TABLES AND FIGURES	69
APPENDIX B: WORD STIMULI.....	96
APPENDIX C: MANIPULATION CHECK QUESTIONS FROM STUDY 2.....	97
APPENDIX D: SAMPLE PASSAGE AND COMPREHENSION QUESTIONS	98

LIST OF TABLES

Table 1. Mean SAM Scores by Block Order.	69
Table 2. Mean SAM Score by Block.	70
Table 3. Paired-sample t-test Results for SAM Scores by Block.	71
Table 4. Mean Accuracy by Block Order.	72
Table 5. Mean Accuracy by Block.	73
Table 6. Paired-sample t-test Results for Accuracy by Block.	74
Table 7. Marginal Means for Comprehension Accuracy by Load Type.	75
Table 8. Marginal Means for Comprehension Accuracy by WMC, Subjective Arousal, and Load Type.	76
Table 9. Marginal Means for WMC, Subjective Arousal, and Load Type.	80
Table 10. Marginal Means for Comprehension Accuracy by Load Type.	81
Table 11. Marginal Means for Comprehension Accuracy by WMC, Subjective Arousal, and Load Type.	82
Table 12. Marginal Means for WMC, Subjective Arousal, and Load Type.	86
Table 13. Means for Accuracy by Ethnicity.	87
Table 14. Mean SAM Score by Block Order.	88
Table 15. Mean SAM Score by Block.	89
Table 16. Paired-sample t-test Results for SAM Score by Block.	90
Table 17. Mean Accuracy by Block Order.	91
Table 18. Mean Accuracy by Block.	92
Table 19. Paired-sample t-test Results for Accuracy by Block.	93
Table 20. Means for Accuracy by Ethnicity.	95

LIST OF FIGURES

Figure 1. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under No Load.....	77
Figure 2. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Neutral Load.....	78
Figure 3. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Arousing Load.....	79
Figure 4. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under No Load.....	83
Figure 5. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Neutral Load.....	84
Figure 6. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Arousing Load.....	85
Figure 7. Marginal Means for Comprehension Accuracy by Load Type.	94

1 INTRODUCTION

Scores from high-stakes tests such as the Scholastic Aptitude Test (SAT) and Test of English as a Foreign Language (TOEFL) are commonly used as criteria for college admission. However, there has been an increase in research showing that standardized testing does not accurately reflect ability, specifically across minority groups such as African Americans, Latinos, and Native Americans (e.g., Scherbaum & Goldstein, 2008; Walpole, Burton, Kanyi, Jackenthal, 2002). Although research has investigated the reliability of standardized tests between different groups, less work has investigated the individual differences that may contribute to poor test performance.

1.1 Working Memory

One area in which susceptibility to failure has been examined is math performance. Beilock (2008) observed that some individuals, despite underlying ability, will “choke” under the pressure of a high-stakes test. Studies have shown that this underperformance is due to an anxiety-induced depletion of cognitive resources, in particular working memory capacity (WMC; e.g., Beilock, Kulp, Holt, & Carr, 2004; Gimmig, Huguet, Caverni, & Cury, 2006). This anxiety-induced reduction in WMC has also been demonstrated in grammatical reasoning tasks (MacLeod & Donnellan, 1993), listening comprehension in second-language learners (Chen & Chang, 2009), and reading comprehension (Calvo & Eysenck, 1996; Rai, Loschky, Harris, Peck, & Cook, 2011).

1.1.1 Working Memory Load

Recently, there has been a growing interest in the working memory (WM) system and its role in test performance. Previous research has consistently demonstrated that a variety of demands (e.g., divided attention, anxiety, and emotional arousal) can act as a “load” on the WM

system, thereby reducing the overall WMC that is available for use in other tasks (e.g., Beilock & DeCaro, 2007; Beilock, Rydell, and McConnell, 2007; Calvo, Ramos, & Estevez, 1992; Mangels, Good, Whiteman, Maniscalco, & Dweck, 2012; Matthews & Campbell, 2010; Sliwinski, Smyth, Hofer, & Stawski, 2006). That is, emotional states like anxious worry and emotional arousal can usurp available WMC away from the current task, leaving less capacity for current task demands (Ilkowska & Engle, 2010). By reducing WMC, cognitive task performance may suffer because the WM resources required to perform these tasks are insufficient. Emotional load effects have been consistently demonstrated in academic settings. For example, Mattarella-Micke and colleagues found that math anxiety acted as a load on WMC to reduce performance on a modular arithmetic task (Mattarella-Micke, Mateo, Kozak, Foster, & Beilock, 2011). Similarly, Chen and Chang (2009) found that foreign language anxiety acted as a cognitive load, thereby reducing WM resources, and leaving fewer resources available for test performance (General English Proficiency Test). Additionally, in a series of studies, Schmader and Johns (2003) investigated the effect of stereotype threat (women and Latinos) on WMC and math test performance. Authors found that stereotype threat caused reduced test performance, and this relationship between stereotype threat and test performance was mediated by a temporary reduction in WMC. Schmader and Johns concluded that induced stereotype threat can act as an additional cognitive load in stigmatized groups (women and Latinos), thereby reducing their available cognitive resources, and resulting in reduced performance rates.

1.2 Arousal

The Yerkes-Dodson Law (1908) states that there is an inverted-u relationship between arousal level and performance, and that arousal and task difficulty have a relationship such that as task difficulty increases, arousal increases. This suggests that there is an optimal level of

arousal for individuals, depending on the task. Indeed, more recent work (e.g., Cassady & Johnson, 2002) supports this inverted-u relationship and shows that a moderate level of arousal may actually benefit performance. Derakshan and Eysenck (2010) proposed that the Yerkes-Dodson Law may partly be explained by Easterbrook's (1959) work on cue utilization. Easterbrook proposed that increasing levels of arousal reduce the scope of attention to task-relevant information, thus improving performance; however, when too aroused, relevant cues may fall outside the scope of attention. Also, level of arousal is influenced by individual differences such as susceptibility to arousal and anxiety. More recently, Schwarzer and Jerusalem (1992) also found that when arousal levels are too high, individuals may be unable to restrict their scope of attention to enough relevant cues or they may restrict their attention to irrelevant cues.

1.3 Current Study

Although research has often shown that WMC load is detrimental to a variety of dual-task applications in contrived laboratory contexts, what needs more empirical attention is determining the extent to which loads on WMC influence how people read and comprehend written information. Specifically, the emotional arousal experienced during testing, especially high-stakes testing, is likely to act as a cognitive load if arousal is too high, and thus work to reduce performance rates in individuals who might otherwise perform quite well. For example, Fartoukh and colleagues found that emotion acts as a cognitive load during writing, as evidenced by more spelling errors and decreased writing fluidity in emotional writing conditions compared to the neutral writing condition (Fartoukh, Chanquoy, & Piolat, 2012). Further, this detrimental effect of emotional arousal may be further compounded in those that are trait anxious, as anxiety

has also been linked to a decrease in WMC (e.g., Beilock & Carr, 2005; Schmader, Johns, & Forbes, 2008).

In the current studies, I tested whether WMC, emotional arousal, and trait anxiety (TA) interact to reduce individuals' ability to comprehend written information. Comprehension was tested using an ecologically valid paradigm (Study 2) set to simulate standardized testing context and procedures.

2 BACKGROUND

2.1 Working Memory Capacity

2.1.1 *Baddeley's Model*

Working memory is a limited-capacity memory system that allows for the temporary storage and manipulation of information (Baddeley & Hitch, 1974). This WM system consists of the central executive, the episodic buffer, the visuo-spatial sketchpad (VSSP), and the phonological loop (PL). The central executive provides attention control of the entire system, whereas the episodic buffer allows communication between components of the WM system and long-term memory, and binds information into chunks (Baddeley, 2003). The VSSP and PL are known as the “slave systems” and specialize in the processing of specific kinds of information. The VSSP is responsible for the storage and manipulation of visual and spatial information; whereas, the PL is devoted primarily to language processing (Baddeley, Gathercole, & Papagno, 1998). The PL can further be divided into the phonological store, which temporarily stores phonological information, and a rehearsal process, which delays the deterioration of information in the phonological store (Baddeley, et al., 1998). Additionally, the rehearsal process is responsible for taking in visual information (i.e., text) and converting it into a phonological code that can be processed by the storage component (Baddeley, 2003; Coltheart, 1993).

In a seminal study, Daneman and Carpenter (1980) found that WMC predicted reading comprehension. Work by McVay and Kane (2012) also supported this finding. Similarly, Christopher and colleagues found that WMC reliably predicted reading comprehension in a sample of fourth and fifth grade children (Christopher et al., 2012). Further, Engle and colleagues found that WMC predicted reading comprehension in first, third, and sixth grade children (Engle, Cantor, & Carullo, 1992). This finding is consistent with capacity theories of

attention (e.g., Kahneman, 1973) as reading requires executive resources such as attention, encoding, storage, and maintenance, although reading becomes highly automated, such that it requires few attention resources (Just & Carpenter, 1992; Waters & Caplan, 1996).

Previous work has revealed several characteristics of the phonological store and rehearsal process. For example, the similarity effect is the finding that words that are phonologically similar (e.g., cat, man, can) result in poor recall, compared to words that are phonologically *dissimilar* (e.g., cow, day, pit; Baddeley, 2003). Similarly, the irrelevant sound effect occurs when irrelevant speech (e.g., hearing an irrelevant conversation) occurs during rehearsal, causing reduced recall as compared to conditions in which no irrelevant speech sounds are heard (Baddeley, 2003). Additionally, the number of words one can maintain in the PL is directly related to the length of the words, with fewer longer words being maintained compared to shorter words (Baddeley, 2003). This effect has been attributed to the finding that one can store as many words as can be rehearsed in approximately two seconds, and since longer words take longer to rehearse, fewer of them can be maintained in the PL at a given time (Baddeley et al., 1975b). However, the word length effect can be prevented by concurrent articulatory rehearsal, and is commonly demonstrated by rehearsing irrelevant verbal information (e.g., repeating the word “the” or counting repeatedly from 1 to 6) while trying to memorize additional verbal information (e.g., a word list; Baddeley, 2003). This effect is known as articulatory suppression, and in addition to reducing the benefit of short words, has also been shown to reduce recall for word lists (Coltheart, 1993).

Articulatory suppression has often been used as a manipulation in reading comprehension studies (e.g., Calvo & Eysenck, 1996; King & Just, 1991). For example, Baddeley and colleagues (1981) examined the effect of articulatory suppression on reading, and found that

those engaged in articulatory suppression had a reduced ability to detect anomalous words and word order errors within sentences (Baddeley, Eldridge, & Lewis, 1981). Similarly, Coltheart and colleagues (1990) found that participants who performed articulatory suppression while reading sentences for acceptability had more false alarms than participants who did not perform articulatory suppression (Coltheart, Avons, & Trollope, 1990). This effect of reduced recall and increased errors is thought to occur because the concurrent articulation is preventing the phonological recoding process, which is necessary to convert text into phonological code that can be processed by the storage component of the PL (Baddeley, 2003; Coltheart, 1993).

However, research is not consistent on whether taxing the PL specifically is detrimental to task performance. Eysenck, Payne, and Derakshan (2005) found that high TA was associated with decrements on a WM task only when a concurrent secondary task involved the use of the central executive component of WM, but not when the secondary task involved the use of the PL. Conversely, other studies (e.g., Derakshan & Eysenck, 1998; MacLeod & Donnellan, 1993) have found that loading the PL is detrimental to dual-task performance; however, in these studies both the primary and secondary tasks were verbal in nature. So, it seems likely that performance is task dependent. That is, generally dual-task performance suffers when both the primary and secondary tasks tax the central executive. However, loading the PL is only detrimental to performance when both the primary and secondary tasks require word processing.

2.1.2 Other Models of Working Memory

I have focused my discussion of WM on Baddeley and Hitch's conceptualization because it is the model most commonly found in the literature, and likely the most influential; however, there are several other theories of WM. I will briefly review some of these. For example, Cowan (1999) described WM as cognitive processes that are in a readily accessible state. This

theory, the Embedded Processes Theory, is centered around a limited-capacity focus of attention that interacts with activated long-term memory representations (Cowan, 1999). In contrast, Engle and colleagues (Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999) have defined WM as the ability to control attention, and consists of short-term memory storage, and, more importantly, attention. According to this theory, attention is used to maintain and process information for the task at hand, while at the same time inhibiting irrelevant or distracting information. Also, Just and Carpenter (1992) proposed a WM model, Capacity Constrained Comprehension Theory, specifically for language processes. Just and Carpenter defined WMC as the maximum number of active representations (e.g., a word, phrase, etc.) in WM for either storage or processing, and individuals differ in how much WMC they have. Both the speed and accuracy of one's language comprehension is determined by how much WMC a person has. Caplan and Waters (1995) also proposed a WM model for language processes, but posited that verbal working memory has subsystems for different kinds of verbal tasks. Specifically, there is a subsystem for determining word or sentence meaning (i.e., interpretive processing), and another subsystem for using the extracted word or sentence meaning in further processing such as converting it to long-term memory or reasoning (i.e., post-interpretive processing). Even though the theories discussed above may seem quite different, they have a common thread among them: the key role of executive processes, specifically attention.

2.1.3 Hypotheses Regarding Working Memory

Although studies to date have consistently found reduced language processing performance when subjects are performing articulatory suppression, research has not examined the effect of taxing the PL with an emotional or arousing WM load specifically. It follows that the emotion induced by the emotional words in the PL act as a "load" on WMC, as it demands

attention resources, to the detriment of reading comprehension. That is, the disturbing nature of the arousing words occupies attention resources beyond just word processing. Moreover, people who are especially sensitive to anxiety provoking stimuli (i.e., TA individuals) should demonstrate the greatest detriment in their performance.

2.2 Cognitive Load Theory

Cognitive load is a process, task, or stimulus that limits the available WMC (i.e., executive attention) needed to process information (Sweller, 1988). Cognitive load increases as demands on attention increase (e.g., via task complexity, dual-task performance, trying to suppress extraneous information), and given that we have a fixed cognitive capacity, cognitive resources devoted to one task decrease the amount of resources available for additional tasks (Sweller, 1988). This is in contrast to automatic processes, which require little or no attention resources, and do not act as a load on WMC (Cohen, Dunbar & McClelland, 1990). In experimental manipulations, cognitive load is often created by requiring participants to switch back and forth between multiple tasks, or by having participants actively maintain a string of letters, words, or numbers in memory while performing another task simultaneously (Lavie, 2010). The cognitive load theory of attention and cognitive control posits that interference from distracters or irrelevant information depends on the level and type of load on cognitive resources such as WMC (Lavie, Hirst, Fockert, & Viding, 2004). Individuals with high cognitive load will be more prone to distraction by irrelevant information, whereas those under a low cognitive load will be less distracted. For example, the classic Baddeley and Hitch (1974) study examined memory for prose passages under 3 WM load conditions: subjects repeated a single number while reading passages, repeated 3 numbers, or repeated 6 numbers. The results revealed that performance (i.e., memory for the passage) was greatly impaired under heavy cognitive load

(i.e., 6 numbers) compared to medium (3 numbers) or low cognitive load (1 number). MacLeod and Donellan (1993) investigated the effects of anxiety on WMC using a load manipulation. Participants of high and low anxiety levels performed a grammatical reasoning task under low cognitive load (subjects kept in mind a string of 6 zeros) or high cognitive load (subjects kept in mind a string of 6 random numbers). The results revealed an interaction between load and anxiety: performance on the grammatical reasoning task was mediated by a reduction in WMC such that those high in TA (HTA) showed a disproportionate reduction in WMC compared to low TA (LTA) individuals.

2.2.1 Cognitive Load and Attention

Recently, research has been focused on the possibility that the relationship between WMC and performance under load is due to differences in the ability to allocate attention to maintain relevant information and suppress irrelevant distracters, with individuals high in WMC (HWMC) showing superior performance compared to individuals low in WMC (LWMC) (e.g., Engle & Kane, 2004). Engle (2002) suggested that HWMC and LWMC persons do not differ in the amount of attention resources (i.e., WMC) they have per se, but differ in terms of how well they can flexibly and efficiently allocate these resources, especially in times of interference or when demands on WMC are high. Similarly, Eysenck, Derakshan, Santos, and Calvo (2007) explained the finding that anxiety and performance are inversely related as a matter of individual differences in the ability to control attention, with HTA persons having an attention bias toward threat-related items (internally or externally generated).

2.2.2 Arousal as a Load

Multiple studies have demonstrated that heightened arousal during a task can act as a load on WMC for some individuals. For example, Mattarella-Micke and colleagues (2011) examined

the effects of physiological stress response (cortisol), trait math anxiety, and WMC on math performance. The results indicated that participants with high anxiety had higher cortisol levels and showed poorer performance compared to participants with low anxiety, who showed higher performance rates when cortisol levels were high. The relationship between cortisol levels and performance also depended on WMC. Specifically, LWMC persons' performance remained constant regardless of cortisol levels and problem difficulty; however, on difficult math problems participants with HWMC and high math anxiety performed significantly worse compared to HWMC persons with low math anxiety when cortisol levels were high. Also, Ashcraft and Kirk (2001) examined the effect of anxiety on dual-task performance. Specifically, individuals with low and high math anxiety performed addition problems while keeping random letter strings in mind. The results suggest that individuals with high math anxiety had significantly more math errors under dual-task performance compared to individuals with low math anxiety. Ashcraft and Kirk concluded that the finding of a differential reduction in accuracy for high versus low anxiety individuals was because high anxiety individuals' emotional reaction to the difficult math problems consumed executive resources that could otherwise have been used for solving math problems.

2.2.3 Hypotheses Regarding Arousal

According to the Cognitive Load Theory and the literature reviewed above, it is likely that heightened emotional arousal will act as a cognitive load to cause a decrement in reading comprehension. One possible mechanism for this relationship is that as emotional arousal increases, attention becomes increasingly divided between the emotional state and the task (reading comprehension, here), leaving less attention and processing resources for the task. As a

result, performance will suffer, especially in those that have a greater susceptibility for emotional arousal (i.e., individuals with TA).

2.3 Trait Anxiety

In testing situations, anxiety is generally discussed in terms of 2 components (Liebert & Morris, 1967). One, emotionality, refers to heightened physiological responses such as increased heart rate and galvanic skin response. Two, worry, refers to self-deprecating ruminations. Worry is typically manifested through internal dialogue; for example, “I’m going to fail”, “I’m not going to get into college” (Deffenbacher, 1980). However, Spielberger (1966) drew a distinction between anxiety as a trait and as a state. State anxiety refers to a temporary, situation-specific reaction that consists of apprehension and heightened autonomic arousal. Trait anxiety is a personality characteristic, and refers to one’s overall likelihood of responding to an event or stimulus with state anxiety. Research has shown that high levels of TA are associated with performance decrements (Darke, 1988a; Mueller & Overcast, 1976). More specifically, anxiety-induced performance deficits have been attributed to a temporary reduction in WMC (Eysenck, 1979). Consistent with Cognitive Load Theory (Sweller, 1988), anxiety has been shown to consume WM resources, leaving fewer resources for dual-tasks, and thus decreasing performance (Shackman, Sarinopoulos, Maxwell, Pizzagalli, Lavric, Davidson, 2006). For example, MacLeod and Donnellan (1993) found an interaction between TA and WM load such that those with high trait anxiety (HTA) had disproportionately reduced WMC and had increased decision latencies on a grammatical reasoning task compared to those with low anxiety (LTA). One explanation for the finding of poor performance in high anxious individuals is Interference Theory, which posits that high anxious persons spend more time attending to task irrelevant information (e.g., worry about test performance and consequences of poor test performance,

comparing one's ability or performance to that of others, etc.), and it is these task-irrelevant thoughts that cause individuals to devote less attention to the task they are performing (Wine, 1980). Similarly, according to Attentional Control Theory (ACT; Eysenck, et al., 2007), anxiety causes an attention impairment such that high anxiety increases the detection of and attention to threat stimuli, with difficulty disengaging from threat stimuli, and decreased attention to tasks that do not contain threat stimuli. That is, highly anxious persons tend to be stimulus-driven, bottom-up processors and often have difficulty maintaining top-down attention control. However, performance impairments are task-dependent whereby anxiety typically does not affect performance effectiveness (e.g., accuracy), but affects processing efficiency (the amount of cognitive resources used given the level of effectiveness).

Similarly, Processing Efficiency Theory (PET; Eysenck & Calvo, 1992) posits that anxiety has both a direct detrimental effect and an indirect compensatory effect. The compensatory effect attempts to counteract detrimental effects by recruiting additional strategies or resources to increase performance. However, when these additional resources cannot be utilized, anxiety impairs performance (i.e., effectiveness). For example, Calvo and Eysenck (1996) investigated the effect of anxiety and compensatory reading strategies on reading comprehension, and found that the use of the PL was more important to those with high anxiety, as their reading comprehension suffered disproportionately compared to those with low anxiety when either component of the PL was disrupted. This would suggest that use of the PL is a strategy or resource that high anxiety persons use to maintain performance levels. Similarly, in a series of studies, Calvo, Eysenck, Ramos, & Jimenez (1994) investigated the role of test anxiety, WM load, and induced state anxiety on the use of reading strategies. Working memory load was induced by having participants read passages while under either articulatory suppression

(continually repeated “ola”), concurrent speech (heard an irrelevant story), and or reading alone. Evaluative stress was induced by advising subjects that their performance was being videotaped, their performance was an indicator of intelligence and academic success, and individual results would be evaluated and compared with those of other participants. This study yielded several important findings. The results showed, first, that neither concurrent speech nor articulatory suppression differentially affected performance as a function of anxiety level (effectiveness), although both of these concurrent tasks significantly reduced reading speed and increased the use of reading regressions and articulatory rehearsal for high anxious participants (efficiency). Second, the effects of anxiety depended on evaluative stress: those with high anxiety were less efficient under stress conditions, whereas, low anxiety persons were less efficient under *no* stress. Performance on the reading span task (a WM measure) failed to predict reading comprehension, but deficits in vocabulary knowledge were found to be partly responsible for the effects of anxiety on reading as high anxious persons had lower vocabulary knowledge compared to low anxious persons.

2.3.1 Hypotheses Regarding Trait Anxiety

Although Calvo, Eysenck, and colleagues (e.g., Calvo & Eysenck, 1996; Calvo, Eysenck, & Estevez, 1994; Calvo, et al., 1994; Eysenck & Byrne, 1994; Eysenck & Calvo, 1992; Eysenck, et al., 2007) have made important contributions to our understanding of factors that affect and mechanisms that underlie reading comprehension, more research in this area is needed as consequences for poor performance on high-stakes tests can be severe (e.g., denied college admission). Given the quality and breath of Calvo, Eysenck, and colleagues’ work, I used this work as a platform for the current study. Specifically, I improved upon the ecological validity of this previous work as pertains to high-stakes testing (outlined below), and further investigated

the role of an emotional cognitive load (as may be seen in high-stakes testing situations) and subjective arousal on reading comprehension performance in a context specific manner. In the current studies I measured reading comprehension using passages and multiple-choice questions from an SAT practice guide; whereas past studies (e.g., Calvo & Eysenck, 1996; Calvo, et al., 1994) have measured comprehension via recognition true/false questions. Additionally, in the current study, I measured comprehension directly after each reading passage as opposed to measuring comprehension at a delay with all testing completed on the same day (Calvo, et al., 1994). Additionally, I examined TA as opposed to test anxiety (e.g., Calvo & Eysenck, 1996; Calvo, et al., 1994), as well as the roles of an emotional load and subjective arousal on WMC limitations during test performance.

Specifically, in Study 1, I expected that an emotional cognitive load would adversely affect reading comprehension beyond decrements seen under general cognitive load, and would mostly affect HTA persons, especially if they also had HWMC. Further, whereas some arousal may benefit LTA persons, arousal was expected to cause decrements in performance for HTA persons. However, this benefit of arousal for LTA persons would be diminished under state anxiety (Study 2).

2.4 Arousing Stimuli

Given the consistent finding of the impact of emotional arousal on cognitive processes, Baddeley (2007) added a hedonic detector component to his existing WM model. He proposed that the hedonic detector acts as an attention filter for the central executive whereby emotional information must reach a certain threshold before it enters the central executive for further processing. Further, this threshold is influenced by genetic and personality factors as well as

situational factors such as mood. Additionally, those high in anxiety chronically have a lower threshold for emotional information compared to those who are low in anxiety.

Whereas research is being conducted to investigate the proposed hedonic detector and factors influencing the processing of emotional stimuli (e.g., Baddeley, 2012), the results are mixed for studies that employ arousing words as stimuli. For example, Bayer, Sommer, and Schacht (2011) investigated task load and emotional words using a pupillary response measure, and found that the arousal of emotional words does not activate the autonomic nervous system, but works on a cognitive level to facilitate word processing. Huang, Baddeley, and Young (2008) investigated the effect of emotional word stimuli on attention control, and found that the emotionality of words only interfered with performance if the emotional words were processed semantically versus phonologically. Authors posited that the emotional words occupied more attention resources compared to neutral words, but only when words were processed semantically. Further, Lindstrom and Bohlin (2011) found that stimulus emotionality facilitates WM performance for both positively and negatively valenced stimuli, and posited that arousal might be driving this effect. Additionally, Naveh-Benjamin and colleagues investigated the effect of emotionally arousing word stimuli on younger and older adults' associative memory (Naveh-Benjamin, Maddox, Jones, Old, & Kilb, 2012). The results indicated that item memory was improved with emotionally arousing words for both younger and older adults; however, associative memory for word pairs was not improved with emotionally arousing words.

Although there is some evidence that emotionally arousing stimuli has a facilitating effect on word processing in some contexts (i.e., recall of emotional words), emotionally arousing stimuli has been shown to reduce dual-task performance. For example, Gotoh (2008) demonstrated that both positive and negative emotional stimuli were likely to capture attention,

causing increased response times for same-different decisions. Fox and colleagues (2001) posited the reason that emotional stimuli can be problematic (i.e., capture attention) is due to difficulty disengaging attention from these stimuli (Fox, Russo, Bowles, & Dutton, 2001), and this effect is especially pronounced in those with high anxiety (Yiend & Mathews, 2001). Similarly, several studies have demonstrated that threat-related words (e.g., death, disease, failure) are more likely to capture attention than neutral words (McNally, Reimann, & Kim, 1990; Segal, Gemar, Truchon, Guirguis, & Horowitz, 1995). For example, Williams and colleagues found that people were slower to name colors of threatening words compared to naming colors of neutral words on the emotional Stroop task (Williams, Matthews, & MacLeod, 1996). Also, MacLeod and colleagues (1986) found that people were faster at detecting target items if the items were displayed in the same location as previously threatening words versus emotionally neutral words (MacLeod, Mathews, & Tata, 1986). Levens and Phelps (2008) used a modified recency-probes paradigm to investigate the role of emotion in resolving conflict between relevant and irrelevant information (i.e., interference resolution) in WM. The results revealed that word stimuli with high emotional arousal (both positively and negatively valenced) aided in making the correct response when relevant and irrelevant information was in conflict in WM compared with neutral word stimuli. Authors posited that emotion or emotional stimuli may differentially influence each component of the WM system.

In the language processing domain, Jimenez-Ortega and colleagues investigated the effects of emotionally arousing texts on error detection (Jimenez-Ortega et al., 2012). Specifically, adult subjects read either a positively, negatively, or neutrally valenced paragraph, and then made judgments of correctness on neutral sentences containing either syntactic or semantic errors, or no errors. The results showed an effect of emotion on task performance, with

increased reaction times and errors in the positive paragraph condition compared to the negative and neutral paragraph conditions. Similarly, Fartoukh and colleagues investigated whether emotion constitutes a cognitive load during writing in a sample of fourth and fifth-graders (Fartoukh, Chanquoy, Piolat, 2012). Participants produced either a positively, negatively, or neutrally valenced text, and the effects of emotion on spelling errors and writing fluidity were examined. The results showed that emotion did, in fact, act as a cognitive load on writing as there were more spelling errors in the emotional writing conditions compared to the neutral writing condition. Additionally, a decrease in writing fluidity was seen in the negative writing condition.

Further, previous work has shown that negative emotional states can reduce available WMC, especially for verbal information (Gray, Braver, & Raichle, 2002; Gray, 2001) as the internal dialogue that accompanies such states is verbal in nature. For example, Beilock, Rydell, and McConnell (2007) investigated whether worry (a negative emotional state) would differentially affect performance on tasks that required more verbal processing versus less verbal processing. Participants completed math problems that varied in the amount of verbal resources needed, while under stereotype threat or not. The results showed that participants that were under stereotype threat performed at a lower rate on math problems that required substantial verbal resources to complete. Authors reasoned that stereotype threat causes an internal dialogue of worries, which is verbal in nature, and these worries take up attention resources that could otherwise be used for task completion.

2.4.1 Hypotheses Regarding Arousing Stimuli

Although extant literature finds that heightened emotionality and arousal act as a load on WMC, reducing processing resources and subsequently performance, research demonstrates a

mix of findings regarding emotionally arousing word stimuli in general. Currently, it is unknown how emotionally arousing words affect the comprehension of text passages. It could be the case that emotional words serve to increase arousal and act as an additional cognitive load. That is, the recall of emotional words may be facilitated, as previous research suggests (Bayer, et al., 2011; Lindstrom & Bohlin, 2011; Naveh-Benjamin, et al., 2012), but at the cost of decreased comprehension of the passage.

2.5 Current Study Overview

Baddeley and Hitch (1974) examined prose recall compared to unrelated sentences and single words, while taxing the PL (with digit strings), and observed that load was an important factor in determining the degree to which memory for prose was affected by divided attention. Additionally, previous research has consistently demonstrated that emotional cognitive load acts on the central executive, reducing WMC, and thus decreasing task performance (e.g., Schmader & Johns, 2003). The emotional arousal accompanying testing, especially high-stakes testing, may act as a cognitive load on WMC and thereby decrease performance. However, studies have not examined the effect of an arousing load that requires word processing as the primary task. That is, previous studies have not differentiated between a general load on comprehension and an arousing load, akin to worry, that may further influence performance. I aimed to fill this gap by using words to induce emotion like that in a high-stakes testing situation that may subsequently affect reading comprehension.

In Study 1, I examined both WMC (high, medium, low) and load type (none, neutral, and arousing) to investigate WMC limitations in test performance. To load the PL, participants rehearsed neutral or arousing words (compared to those who had no load) while reading several passages and answering comprehension questions about each passage. Additionally, TA and

subjective arousal were examined as these factors may differentially affect reading comprehension as a function of WM and cognitive load. Study 2 further explored the relationship between the above variables under context-specific anxiety.

3 EXPERIMENT 1

The purpose for the current study was to test whether the type of cognitive load (emotional or neutral) differentially influences reading comprehension, indexed as test performance, as a function of TA, subjective arousal, and WMC. I hypothesized that LWMC and HWMC individuals would perform similarly under no load. However, when loaded, LWMC individuals would perform poorly compared to the no-load condition and compared to HWMC persons in all conditions. High WMC persons were expected to demonstrate similar performance across conditions. In addition, I hypothesized that emotional cognitive load would adversely affect reading comprehension beyond decrements seen under general cognitive load (i.e., neutral load), especially in HTA persons that are also HWMC. Further, while some subjective arousal may benefit LTA persons, subjective arousal was predicted to cause decrements in performance for HTA persons.

3.1 Methods

3.1.1 *Participants*

One-hundred seventy-five participants (136 women) were recruited from the Georgia State University Psychology Department's subject pool. All participants were over the age of 18 and the sample consisted of 76 African-Americans, 35 Caucasians, 26 Asians, 14 Hispanics, and 24 individuals from other racial backgrounds. Participants received course credit for participation.

3.1.2 *Materials*

3.1.2.1 Word Stimuli

For the arousing cognitive load condition, I selected 15 negatively valenced, high arousal words from the Affective Norms for English Words (ANEW; Bradley & Lang, 1999) (See

Appendix A for full list of word stimuli). I chose the words with the highest arousal ratings (6.00+ on a 9.00 scale) and that were also negatively valenced. All words were bi-syllabic and 5 to 6 letters in length. For the neutral cognitive load condition, I selected 15 neutral words that were length (also bi-syllabic and 5 to 6 letters), familiarity, and frequency-matched to the negatively valenced, high arousal words using ratings from the MRC Psycholinguistic database (Wilson, 1988).

3.1.2.2 State-trait anxiety inventory

The State-Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) is used to assess both state and trait anxiety. The state portion of the test contains 20 statements, and participants indicate to what degree (a 4-point Likert scale) that statement describes how they are feeling at *this very moment* (e.g., I feel calm). The trait portion of the test contains 20 statements in which participants indicate to what degree (also a 4-point Likert scale) that statement describes how they *generally feel* (e.g., I make decisions easily). Scores on each portion (state and trait) range from 20 to 80, with a higher score indicating a greater level of anxiety. This assessment takes approximately 10 minutes to complete.

3.1.2.3 Passages and comprehension questions

A total of 10 passages (1 practice, 9 experimental) with corresponding comprehension questions (2 to 4 each) were taken from an SAT practice guide (Gruber, 2011). Each passage was approximately 150 words in length and covered a variety of subjects (e.g., pyramids, theatre, and history). Comprehension questions were multiple-choice format with five answer choices each (See Appendix C for a sample passage and questions). The presentation of passages was randomized.

3.1.2.4 Self-assessment manikin

The Self-Assessment Manikin (SAM; Lang, 1980) is used to assess the pleasure, arousal, and dominance associated with an emotional reaction to stimuli or events. Each of these factors is represented by 5 person-like figures with corresponding numbers (1 to 9) so the assessment functions as a Likert scale. The figures for the pleasure scale range from smiling and happy to frowning and unhappy. Whereas, the figures for the arousal scale range from excited and wide-eyed to relaxed and sleepy. The figures for the dominance scale show changes in control, and range in size from small to large, with the large figure representing maximum control of the situation, and the small figure representing no control of the situation. Participants indicate which picture most closely corresponds with how they are feeling at that moment. Here, the arousal dimension of the SAM was used to determine subjective arousal, with a rating of 1 indicating a feeling of calm or quiet, and a rating of 9 indicating a feeling of elation, excitement, or being upset. Specifically, a change score was calculated by subtracting the SAM arousal rating for the neutral load block from the SAM arousal rating for the arousing load block. If participants increased in absolute value from the neutral load block to the arousing load block, participants were considered aroused; if participants decreased in subjective arousal or stayed at the same level of arousal, participants were considered not aroused.

3.1.2.5 Automated running memory span task

The automated running memory span task (Run Span; Broadway & Engle, 2010) is a computerized assessment of WMC. A string of 3 to 8 letters is displayed, one letter at a time, each for a duration of 2500 ms. Participants are notified at the beginning of each trial how many letters they will be asked to report (but not the number of letters to be displayed). Participants are required to remember these letters until prompted to recall n of the letters (3 to 6), in

presentation order, via mouse click. There are a total of 24 trials, which vary in difficulty (i.e., the number of letters to be recalled). The Run Span score is the number of correctly recalled letter sequences across all trials. A higher score indicates higher WMC. Score range is 0 to 75.

3.1.2.6 Nelson-Denny

It is necessary to control for language ability given the observation that vocabulary knowledge is positively correlated with reading comprehension (Baddeley, Logie, Nimmo-Smith, & Brereton, 1985), and, additionally, that individuals high in test anxiety show poorer vocabulary knowledge compared to individuals who are low in test anxiety (Calvo, Ramos, & Estevez, 1992). The Nelson-Denny (Brown, Vick-Fishco, & Hanna, 1993) is a 2-part paper and pencil reading assessment for adults. However, only Part I, the vocabulary section, was utilized in this study. The purpose of this subtest is to assess subjects' vocabulary ability, and consists of a multiple-choice synonym test: 80 items, each with 5 answer choices. This section has a time limit of 15 minutes. A higher number of correct responses reflect higher vocabulary knowledge.

3.1.3 Procedure

Participants were tested in groups of 1 to 6 via computer (E-Prime 2.0.08). Participants first completed the Run Span, and then the STAI. Next, participants completed reading passages and comprehension questions. The experiment was blocked by load type (i.e., no load, neutral load, arousing load), and block order was counterbalanced. Participants completed 1 practice passage with comprehension questions prior to completing the 3 experimental blocks.

3.1.3.1 No Load Block

Participants read a series of 3 passages while keeping in mind a string of 5 Xs (i.e., XXXXX). After each passage, participants were asked to answer multi-choice comprehension questions (2 to 4, depending on the passage) pertaining to the passage they just read, using the

computer keyboard. After answering comprehension questions for each passage in the series, participants typed the information they had been keeping in mind (i.e., 5 Xs). After completing the last question for the third and final passage in this block, participants completed the SAM.

3.1.3.2 Neutral Load Block

Participants read a series of 3 short passages as above, but in this series, they were asked to keep 5 neutral words (e.g., logic, candle, assist) in mind while they read each passage. As above, after each passage, participants answered comprehension questions. After the final comprehension question for each passage in the series, participants typed the neutral words they remembered (any order was accepted). After recalling the neutral words for the third and final passage in this series, participants again completed the SAM.

3.1.3.3 Arousing Load Block

Participants read a series of 3 short passages as above, but in this series of passages, they were asked to keep 5 arousing, negatively valenced words (e.g., demon, cancer, brutal) in mind while they read each passage. As above, after each passage in the series, participants answered comprehension questions. After the final comprehension question for each passage in the series, participants typed the arousing words they remembered (any order was accepted). After recalling the arousing words for the third and final passage in this series, participants completed the SAM.

3.1.3.4 Demographic Information

Next, participants were asked for demographic information: gender and ethnicity, indicated via button press. Lastly, participants completed the Nelson-Denny. Participants were then debriefed, thanked for their participation, and released.

3.2 Study 1 Results

3.2.1 Data Preparation and Variable Calculation

Following the practice of previous studies, levels of TA and WMC were each determined via tertile split of the raw scores (e.g., Cassady & Johnson, 2002; Derakshan, Smyth, & Eysenck, 2009; Minnaert, 2003). To determine subjective arousal, a change score was calculated by subtracting the SAM arousal rating for the neutral load block from the SAM arousal rating for the arousing load block. If participants increased in absolute value from the neutral load block to the arousing load block, participants were considered subjectively aroused; if participants decreased in subjective arousal or stayed at the same level of arousal, participants were considered not subjectively aroused.

Outliers beyond 3 standard deviations above and below the mean for each variable were removed. To ensure that participants were actively engaging in the cognitive load manipulation (maintaining neutral or arousing words during reading comprehension task), only trials in which subjects correctly recalled a minimum of 3 of the 5 words after each passage were included in subsequent analyses.

During data preparation, accuracy and subjective arousal were examined by block to verify that there were no order effects present. There were 6 possible block combinations, and means for each block combination were inspected. When inspecting subjective arousal means, in 5 of the 6 block orders, subjective arousal was greatest in the very last block (See Table 1). This suggests that participants might have increased in subjective arousal, perhaps due to being frustrated by the length of the study. To investigate this possibility further, paired-sample *t*-tests with Bonferroni correction were conducted on SAM ratings by block (See Table 2), and the results revealed that subjective arousal was not significantly different in the first block ($M =$

3.30, $SD = 1.93$) compared to the second block ($M = 3.46$, $SD = 2.09$), $t(174) = -1.14$, $p > .05$ (See Table 3). Likewise, subjective arousal was not significantly different in the second block ($M = 3.46$, $SD = 2.09$) compared to the third block ($M = 3.68$, $SD = 2.17$), $t(174) = -1.74$, $p > .05$. However, subjective arousal was significantly higher in the third (i.e., final) block ($M = 3.68$, $SD = 2.17$) compared to the first block ($M = 3.30$, $SD = 1.93$), $t(174) = -2.55$, $p < .05$. These analyses confirm that subjective arousal increased significantly over the duration of the experiment. It is likely that this effect impacted the calculation of whether participants were considered subjectively aroused or not.

Next, I inspected accuracy rates for the different block orders and discovered that accuracy also appeared to decrease systematically with study length (See Table 4). That is, all participants performed best in the first block, decreased in accuracy in the second block, with a further decrement in the last block. This pattern occurred for all 6 block orders. This suggests that participants' performance may have decreased as a function of study length due to boredom or cognitive fatigue. Although this is inconsistent with the arousal ratings, in which I would have expected reduced subjective arousal as a function of boredom. To investigate this possibility further, I conducted paired-sample t -tests with Bonferroni correction on accuracy of each block (See Table 5), and the results revealed that accuracy was significantly lower in the third block (i.e., last block; $M = .44$, $SD = .19$) compared to both the second block ($M = .54$, $SD = .18$), $t(174) = 6.63$, $p < .01$, and the first block ($M = .67$, $SD = .18$), $t(174) = 14.38$, $p < .01$ (See Table 6). Also, accuracy was significantly lower in the second block ($M = .54$, $SD = .18$) compared to the first block ($M = .67$, $SD = .18$), $t(174) = 8.92$, $p < .01$. These analyses confirm that performance (i.e., accuracy) systematically decreased as a function of study length, and these effects have the potential to obscure any true patterns in the data.

3.2.2 Analysis of Variance

3.2.2.1 Load Type as a Between-Subjects Variable

Given the finding that accuracy decreased systematically with study length, I analyzed data from the first block only using a between-groups design. In order to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension as a function of WMC, TA, and subjective arousal, I conducted a 3 (Load: emotional word, neutral word, no load) X 3 (WMC: high, medium, low) X 3 (TA: high, medium, low) X 2 (Subjective arousal: aroused, not aroused) fully between-groups ANOVA. Vocabulary knowledge ($M = 71.63$, $SD = 13.80$) was entered as a covariate. The dependent variable was accuracy on comprehension questions.

The results revealed a trending main effect of load type, $F(2, 124) = 2.64$, $p = .08$, $\eta^2 = .04$. No main effect of TA, $F(2, 124) = .003$, $p > .05$, WMC, $F(2, 124) = .59$, $p > .05$, or subjective arousal, $F(1, 124) = 1.18$, $p > .05$ was observed. To test whether emotional cognitive load (i.e., arousing words) would adversely affect reading comprehension beyond decrements seen under general cognitive load (i.e., neutral word load), simple contrast analyses with Bonferroni correction were conducted. This analysis showed that comprehension accuracy, $F(2, 124) = 3.79$, $p < .05$, $SE = .04$ was significantly lower in the arousing load ($M = .60$, $SE = .03$) condition compared to both the neutral load ($M = .70$, $SE = .03$) and no load ($M = .70$, $SE = .03$) conditions (See Table 7).

3.2.3 Additional Analyses

3.2.3.1 Tertile Split of WMC without MWMC Group

To investigate the roles of WMC and subjective arousal further, I conducted an additional analysis excluding the MWMC persons group and the variable TA. That is, I conducted a 3 (Load: emotional word, neutral word, no load) X 2 (WMC: high, low) X 2 (Subjective arousal: aroused, not aroused) fully between-groups ANOVA in order to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension as a function of WMC and subjective arousal. Vocabulary knowledge ($M = 71.63$, $SD = 13.80$) was entered as a covariate. The dependent variable was accuracy on comprehension questions.

The results revealed a trending 3-way interaction: Load type x Subjective arousal x WMC, $F(2, 100) = 2.58$, $p = .08$, $\eta^2 = .05$. No significant main effect of WMC, $F(1, 100) = .006$, $p > .05$, subjective arousal, $F(1, 100) = .50$, $p > .05$, or load type, $F(2, 100) = 2.27$, $p > .05$ was observed. Follow-up simple contrast analyses with Bonferroni correction were conducted; however, none of these contrasts were found to be significant, $F(2, 100) = 2.27$, $p > .05$. However, the contrast between no load ($M = .70$, $SE = .03$) and arousing load conditions ($M = .62$, $SE = .03$) was trending, $p = .08$, as well as the contrast between the neutral load ($M = .70$, $SE = .03$) and arousing load conditions ($M = .62$, $SE = .03$), $p = .07$.

Despite the finding of non-significant contrasts, I inspected means of WMC, subjective arousal, and load type for patterns (See Table 8). Under no load, HWMC and LWMC persons performed at about the same rate, and maintained this level of performance regardless of subjective arousal (See Figure 1). However, under dual-task conditions, performance depended on whether participants were subjectively aroused or not. Under neutral load, LWMC persons performed at a lower rate if they were subjectively aroused versus not subjectively aroused; whereas, HWMC persons maintained performance rates regardless of subjective arousal (See

Figure 2). However, under an arousing load, LWMC persons increased in accuracy when they were subjectively aroused compared to LWMC persons that were not subjectively aroused. High WM persons showed decreased accuracy rates if they were subjectively aroused compared to HWMC persons who were not subjectively aroused (See Figure 3).

However, to determine whether the relationship between WMC, subjective arousal, and load type was meaningful, I inspected the marginal means and standard error estimates for these variables (See Table 9). Given that the standard error estimates for WMC and subjective arousal were large compared their small mean differences, I concluded that differences in these variables were not reliable, and were not further interpreted. However, mean differences between load type, specifically between the no load and arousing load conditions, and neutral load and arousing load were sizable. Indeed, these differences between load types were found to be trending in the simple contrasts discussed above, and, thus, appear to be interpretable. That is, reading comprehension accuracy was lower under arousing load conditions compared to neutral or no load conditions.

3.2.3.2 Tertile Split of TA without MTA Group

To investigate the roles of TA and subjective arousal further, I conducted an additional analysis excluding the MTA persons group and the variable WMC. That is, I conducted a 3 (Load: emotional word, neutral word, no load) X 2 (TA: high, low) X 2 (Subjective arousal: aroused, not aroused) fully between-groups ANOVA in order to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension as a function of TA and subjective arousal. Vocabulary knowledge ($M = 71.63$, $SD = 13.80$) was entered as a covariate. The dependent variable was accuracy on comprehension questions.

The results revealed a significant main effect of load type, $F(2, 100) = 3.63, p < .05, \eta^2 = .07$. No significant main effect of TA, $F(1, 100) = .03, p > .05$ or subjective arousal, $F(1, 100) = .81, p > .05$ was observed. To test whether emotional cognitive load (i.e., arousing words) adversely affected reading comprehension beyond decrements seen under general cognitive load (i.e., neutral word load), simple contrast analyses with Bonferroni correction were conducted. This analysis showed a significant difference, $F(2, 100) = 3.63, p < .05, SE = .04, \eta^2 = .07$ specifically between the neutral load ($M = .72, SE = .03$) and arousing load ($M = .61, SE = .03$) conditions, $p < .05$ (See Table 10). Additionally, the difference between the no load ($M = .70, SD = .03$) and arousing load ($M = .61, SE = .03$) conditions was trending, $p = .08$. However, the difference between the no load ($M = .70, SD = .03$) and neutral load ($M = .72, SE = .03$) conditions was not significant, $p > .05$.

3.2.3.3 Load as a Within Subjects Variable

The following analysis was originally planned as the main analysis; however, given the observed fatigue effects, I conducted between-groups analyses on the first block of data (above). However, for the sake of exploring the data further, I included the original analysis.

In order to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension, indexed as test performance, as a function of WMC, TA, and subjective arousal, a 3 (Load: emotional word, neutral word, no load) X 3 (WMC: high, medium, low) X 3 (TA: high, medium, low) X 2 (Subjective arousal: aroused, not aroused) mixed-measures ANOVA was conducted with Load as a within-subject variable and WMC, TA, and subjective arousal as between-groups variables. Vocabulary knowledge ($M = 71.63, SD = 13.80$) was entered as a covariate. The dependent variable was accuracy on comprehension questions.

The results revealed no main effect of load, $F(2, 312) = 1.45, p > .05$, subjective arousal, $F(2, 312) = .33, p > .05$ or WMC, $F(4, 312) = .55, p > .05$. However, a significant 3-way interaction: Load x Subjective arousal x WMC, $F(4, 312) = 2.60, p < .05, \eta^2 = .03$ was observed. Trait anxiety was not found to impact test performance significantly, and thus, no further analyses of TA were conducted.

I hypothesized that LWMC and HWMC individuals would perform similarly under no load; however, when loaded, LWMC individuals would perform poorly compared to the no load condition and compared to HWMC persons in all conditions. Lastly, I predicted that HWMC persons would demonstrate similar performance across conditions. To test this, follow-up simple contrast analyses with Bonferroni correction were conducted. However, this analysis yielded no significant contrasts, $F(2, 156) = .29, p > .05$.

Despite the finding of nonsignificant contrasts, I inspected means of WMC, subjective arousal, and load type for patterns. Under no load and when not subjectively aroused, HWMC persons outperformed LWMC persons (See Table 11). If LWMC persons were subjectively aroused, performance increased compared to non-aroused LWMC persons; conversely, HWMC persons' performance decreased (See Figure 4) if they were subjectively aroused versus not aroused. However, under neutral load and when not aroused, LWMC and HWMC persons performed similarly; but, subjectively aroused HWMC persons' performance increased compared to their counterparts who were not subjectively aroused; whereas, LWMC persons decreased in accuracy if they were aroused compared to LWMC persons that were not aroused (See Figure 5). That is, under no load, subjective arousal was beneficial for LWMC persons (but not HWMC persons); however, this benefit disappears for LWMC persons under divided attention (neutral load), but the HWMC persons benefitted from subjective arousal. Under an

arousing load, LWMC persons maintained performance regardless of subjective arousal; whereas, HWMC persons had poorer accuracy if they were subjectively aroused compared to HWMC persons that were not subjectively aroused (See Figure 6).

However, to determine whether the relationship between WMC, subjective arousal, and load type was meaningful, I inspected the marginal means and standard error estimates for these variables (See Table 12). Given that the standard error estimates for WMC and subjective arousal were large compared their small mean differences, I concluded that differences in these variables were not reliable, and were not further interpreted. However, mean differences between load type, specifically between no load and arousing load conditions, and neutral load and arousing load were sizable, relative to standard error. Accordingly, I conducted paired-sample *t*-tests on accuracy rates of the different load types. The results indicate that accuracy was significantly lower in the arousing load condition ($M = .51$, $SD = .21$) compared to both the neutral load ($M = .56$, $SD = .20$), $t(174) = 2.36$, $p < .05$, and the no load conditions ($M = .57$, $SD = .19$), $t(174) = 3.42$, $p < .01$. However, accuracy was not significantly different in the no load condition compared to the neutral load condition, $t(174) = 1.09$, $p > .05$.

3.2.3.4 Accuracy as a Function of Race/Ethnicity

Given previous research demonstrating differential performance between racial/ethnic groups (e.g., Scherbaum & Goldstein, 2008; Walpole, Burton, Kanyi, Jackenthal, 2002), I conducted a 1-way (ethnicity: Caucasian, African-American, Hispanic, Asian, Other) ANOVA on reading comprehension accuracy. The results show a significant main effect of group, $F(4, 170) = 3.65$, $p < .01$, $\eta^2 = .08$. Follow-up simple contrasts show that Caucasians had significantly higher comprehension accuracy ($M = .72$, $SE = .03$) compared to both African-Americans ($M = .64$, $SE = .02$), $p < .01$, and Asians ($M = .64$, $SE = .03$), $p < .01$ (See Table 13).

No other contrasts were significant. However, when vocabulary knowledge was entered as a covariate, the main effect of group dropped out of significance, $F(4, 169) = 1.72, p > .05$.

Further, no simple contrasts were found to be significant. Given that all the above main analyses were conducted with vocabulary knowledge as a covariate, I concluded that results were not biased by any effects of race/ethnicity.

3.3 Discussion

The purpose for the current study was to test whether the type of cognitive load (emotional or neutral) would influence reading comprehension and whether TA, WMC, and subjective arousal would interact with load type to further affect performance. Several analyses were conducted. It was observed that accuracy systematically decreased with study length, which was likely the result of cognitive fatigue or boredom. Accordingly, only the first block of data was analyzed using a between-groups design. Two main findings resulted. Two analyses showed an effect of load type: one analysis showed that performance under an arousing load was significantly lower than performance under a neutral load or no load; the second analysis showed that performance was significantly lower under an arousing load compared to a neutral load. The finding of differential performance under neutral versus arousing load suggests that performance under load depended on the content of the load (emotional load vs. neutral load) such that performance under an arousing load was consistently worse compared to performance under a neutral (i.e., non-emotional) load. The second main finding was that subjective arousal interacted with WMC and load type to influence reading comprehension performance. However, further inspection of this interaction (in both analyses) revealed that no main effects or contrasts were significant. This is likely due to an issue of low power as an initial power analysis revealed that approximately 270 participants would be needed to achieve 80% power; however, the

current analyses contained only 175, and yielded a power of only 66%. Additionally, strong conclusions about WMC and subjective arousal cannot be based on these analyses as high standard error rates were associated with small mean differences on these variables. However, mean differences between accuracy by load type were sizable with relatively small standard error estimates, and, are thus interpretable: accuracy under an arousing load was lower than accuracy under a neutral load or no load. Overall, these findings tentatively, given the results are in the predicted directions, suggest that the type of cognitive load differentially influences reading comprehension.

Although differential test performance was shown as a function of load type, the roles of TA, WMC, and arousal are less clear. It is possible that the observed cognitive fatigue effects obscured the roles of TA, WMC, and subjective arousal in reading comprehension performance; however, this possibility was ruled out by conducting analyses on the first block of data only. Also, low statistical power is likely to have influenced the detection of such effects. Additional possibilities for the lack of differences are discussed in the General Discussion.

The current study was a lab-based study meant as a stepping stone for the more face valid, Study 2. Though the current study was underpowered, the finding of differences between load types gave a nice platform for the second study where state anxiety was induced in a context specific manner.

4 EXPERIMENT 2

Given the observation in the literature that individual differences in TA and WMC most frequently appear under stress (e.g., Berggren, & Derakshan, 2013; Calvo, et al., 1994; Eysenck, et al., 2007; Ilkowska & Engle, 2010), I induced state anxiety in a context-specific manner to investigate further the roles of WM, subjective arousal, and TA on reading comprehension performance.

I hypothesized that an emotional cognitive load would adversely affect reading comprehension beyond decrements seen under general cognitive load as seen in Study 1. I expected LWMC persons to perform best under no load, but worse than HWMC persons in all conditions except for the case of HWMC persons that were also HTA. High WM persons were expected to maintain performance across load conditions. No benefit of subjective arousal for LTA persons (as predicted in Study 1) was expected under state anxiety.

4.1 Methods

4.1.1 *Participants*

Three-hundred sixty-five participants (284 women) were recruited from the Georgia State University Psychology Department's subject pool. All participants were over the age of 18, and the sample consisted of 187 African-Americans, 80 Caucasians, 40 Asians, 26 Hispanics, and 32 individuals from other racial backgrounds. Participants received course credit for participation.

4.1.2 *Materials*

All materials were the same as Study 1.

4.1.3 Procedure

The procedure was the same as in Study 1. In addition, I induced state anxiety by advising participants that their performance was an indication of intelligence and academic success, that the reading passages were timed (passages were not *actually* timed), that the reading comprehension task was computer adaptive (for every question answered incorrectly, more questions would be added), and that their performance was being videotaped. Previous studies have used these techniques to induce state anxiety (e.g., Beilock & Carr, 2005; Calvo, et al., 1994). A baseline SAM was also added so that participants completed a total of 4 SAMs: once after the STAI and one after each reading block. Additionally, at the end of the reading comprehension section, participants answered 6 manipulation-check questions (see Appendix B for questions and ratings for each question) regarding how bothered they were by the experiment instructions (e.g., performance was videotaped, performance was an indicator of intelligence). Each question was rated on a 9-point Likert scale with a rating of 9 indicating the instruction was highly bothersome and a rating of 1 indicating the instruction was not bothersome at all.

4.2 Study 2 Results

4.2.1 Data Preparation and Variable Calculation

The data were prepared as described in Study 1. To note, although a baseline SAM measure was added to Study 2, to be consistent with Study 1, subjective arousal was calculated by subtracting the SAM arousal rating for the neutral load block from the SAM arousal rating for the arousing load block. If participants increased in absolute value from the neutral load block to the arousing load block, participants were considered subjectively aroused; if participants decreased in subjective arousal or stayed at the same level of arousal, participants were considered not subjectively aroused.

As in Study 1, during data preparation, accuracy and subjective arousal were examined by block to verify that there were no obvious order effects present. There were 6 possible block combinations, and means for each order combination were inspected. Subjective arousal was highest in the last block (4 of 6 orders) except when the no load condition was last (see Table 14). To rule out a method effect, paired-sample *t*-tests with Bonferroni correction were conducted on SAM ratings for each block (See Table 15).

The results revealed that the baseline SAM was not significantly different ($M = 3.43$, $SD = 1.92$) than the first block ($M = 3.49$, $SD = 1.97$), $t(364) = -.67$, $p > .05$ (See Table 16). However, the baseline SAM was significantly lower than both the second block ($M = 3.71$, $SD = 2.09$), $t(364) = -2.73$, $p < .05$, and the third block ($M = 3.76$, $SD = 1.96$), $t(364) = -3.30$, $p < .01$. Further, SAM ratings were significantly higher in both the second block ($M = 3.71$, $SD = 2.09$), $t(364) = -2.63$, $p < .05$ and the third block ($M = 3.76$, $SD = 1.96$), $t(364) = -3.40$, $p < .01$ compared to the first block ($M = 3.49$, $SD = 1.97$). However, SAM rating were not significantly different between the second ($M = 3.71$, $SD = 2.09$) and third blocks ($M = 3.76$, $SD = 1.96$), $t(364) = -.55$, $p > .05$.

From these results, it appears that arousal did increase significantly with study length as subjective arousal was significantly higher in the second and third blocks compared to the baseline SAM measure, and also increased significantly from Blocks 1 to 2. However, it also appears likely that the testing instructions (e.g., videotaping, timed passages, etc.) did not systematically contribute to an increase in subjective arousal as evidenced by the lack of difference in SAM ratings between the baseline SAM measure and the first block SAM measure. That is, the testing instructions were given between the baseline and first block SAM measures; however, these measures were not significantly different from each other. It is likely that this

effect impacted the calculation of whether participants were considered subjectively aroused or not.

From an examination of mean accuracy for each order combination of blocks, accuracy did not decline systematically as a function of study length as was observed in Study 1 (see Table 17). Paired sample t -tests with Bonferroni correction were conducted on accuracy for each block (See Table 18). Results revealed that accuracy was not significantly different in the first block ($M = .57, SD = .20$) compared to the second block ($M = .56, SD = .21$), $t(364) = .40, p > .05$, or the third block ($M = .56, SD = .20$), $t(364) = .78, p > .05$ (See Table 19). Likewise, accuracy was not significantly different in the second block ($M = .56, SD = .21$) compared to the third block ($M = .56, SD = .21$), $t(364) = .31, p > .05$. These results confirm that accuracy did not decline as a function of study length.

4.2.2 Analysis of Variance

In order to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension, indexed as test performance, as a function of WMC, TA, and subjective arousal, a 3 (Load: emotional word, neutral word, no load) X 3 (WMC: high, medium, low) X 3 (TA: high, medium, low) X 2 (Subjective Arousal: aroused, not aroused) mixed measures ANOVA was conducted with Load as a within-subject variable and WMC, TA, and subjective arousal as between-subject variables. Vocabulary knowledge ($M = 72.58, SD = 12.94$) was entered as a covariate. The dependent variable was accuracy on comprehension questions as in Study 1.

The results revealed a main effect of load type, $F(2, 692) = 11.29, p < .001, \eta^2 = .03$. Inspection of marginal means shows superior performance under no load with a reduction in accuracy under neutral load and a further decrement under an arousing load for all participants

(See Figure 7). Follow-up paired-sample t -tests (with Bonferroni correction), $t = 6.41$, $p < .001$ confirmed that comprehension accuracy was significantly lower in the arousing load condition ($M = .47$, $SD = .19$) compared to the neutral load condition ($M = .53$, $SD = .17$) and the no load condition, $t = 21.99$, $p < .001$ ($M = .70$, $SD = .17$). An additional paired-sample t -test, $t = 17.62$, $p < .001$ showed that comprehension accuracy was significantly lower in the neutral load condition compared to the no load condition.

Although I predicted that individual differences in WMC, TA, and subjective arousal would influence test performance, none of these factors significantly affected reading comprehension accuracy, and thus were not further analyzed.

4.2.3 Additional Analyses

4.2.3.1 Accuracy as a Function of Race/Ethnicity

Given previous research demonstrating differential performance between racial/ethnic groups (e.g., Scherbaum & Goldstein, 2008; Walpole, Burton, Kanyi, Jackenthal, 2002), I conducted a 1-way (ethnicity: Caucasian, African-American, Hispanic, Asian, Other) ANOVA on reading comprehension accuracy. The results showed a significant main effect of group, $F(4, 360) = 2.85$, $p < .05$, $\eta^2 = .03$. Follow-up simple contrasts show that Caucasians had significantly higher comprehension accuracy ($M = .62$, $SE = .02$) compared to Asians ($M = .50$, $SE = .03$), $p < .01$. Hispanics ($M = .60$, $SD = .04$) also had significantly higher comprehension accuracy than Asians ($M = .50$, $SE = .03$), $p < .05$ (See Table 20). Additionally, there was a trending difference in accuracy between Caucasians ($M = .62$, $SE = .02$) and African-Americans ($M = .56$, $SE = .01$), $p = .06$. However, when vocabulary knowledge was entered as a covariate, the main effect of group dropped out of significance, $F(4, 359) = .68$, $p > .05$. Further, no simple contrasts were found to be significant when controlling for vocabulary ability. Given that all the

above main analyses were conducted with vocabulary knowledge as a covariate, I concluded that results were not biased by any effects of race/ethnicity.

4.2.3.2 Comparison of accuracy by load type for Studies 1 and 2.

To understand better the impact of the high stakes testing instruction on performance, I compared accuracy in the different load conditions across studies. For Study 1, I again used only data from the first block, and from Study 2, I used a number of observations equal to that used in Study 1. The results from independent t -tests show that accuracy under no load, $t(62) = .34, p > .05$ was not significantly different in Study 1 ($M = .70, SD = .15$) compared to Study 2 ($M = .69, SD = .19$). However, accuracy under a neutral load, $t(62) = 7.03, p < .001$ was significantly lower in Study 2 ($M = .47, SD = .17$) compared to Study 1 ($M = .68, SD = .17$). Also, accuracy under an arousing load, $t(48) = 4.99, p < .001$ was significantly lower in Study 2 ($M = .41, SD = .18$) compared to Study 1 ($M = .62, SD = .20$).

4.2.3.3 Comparison of SAM scores by load type for Studies 1 and 2.

To explore the subjective arousal levels between Studies 1 and 2 further, I compared SAM scores in the different load conditions across studies. For Study 1, I again used only data from the first block, and from Study 2, I used a number of observations equal to that used in Study 1. The results from independent t -tests show that subjective arousal was not significantly different under no load, $t(62) = 1.27, p > .05$ in Study 1 ($M = 3.59, SD = 1.89$) compared to Study 2 ($M = 3.14, SD = 1.86$). Similarly, subjective arousal was not significantly different under neutral load, $t(62) = -.50, p > .05$ in Study 1 ($M = 3.29, SD = 2.08$) compared to Study 2 ($M = 3.46, SD = 1.97$). However, subjective arousal was significantly higher under an arousing load in Study 2 ($M = 3.78, SD = 1.77$) compared to Study 1 ($M = 3.13, SD = 1.82$).

4.3 Discussion

The purpose for the current study was to test whether the type of cognitive load would influence reading comprehension and whether TA, WMC, and subjective arousal would interact with load type to further affect performance under induced state anxiety. Support was shown for my hypothesis that arousing cognitive load would cause a decrement in performance above and beyond that of a general (neutral) cognitive load. This was evidenced by the observation that performance under an arousing load was significantly lower than performance under neutral load and no load. This effect of load type was observed regardless of differences in WMC, TA, or subjective arousal.

Overall, analyses showed that subjective arousal increased as a result of study length as in Study 1; however, no significant increase in arousal was observed after the high-stakes testing instructions. So, it is likely that the testing instructions did not serve to increase subjective arousal. However, it is possible that the testing instructions did serve to increase motivation as accuracy did not decline with study length as observed as in Study 1. It could be that arousal does not impact the accuracy of performance. However, it is more likely, because this was a lab-based experiment with no real-life consequences for poor performance, participants were not aroused *enough* for arousal to impact performance negatively.

The finding of lower performance rates in the arousing load compared to neutral and no load conditions is consistent with the results of Study 1, as well as a lack of finding for effects of TA, WMC, and subjective arousal. In light of the absence of these effects in Study 2, it might be the case that Study 1 would not have yielded significant effects of TA, WMC, or subjective arousal even if more participants had been tested. See the General Discussion for more on the lack of differences for TA, WMC, and subjective arousal.

Additionally, accuracy under no load was similar across studies; however, accuracy was significantly lower under load conditions in Study 2 compared to Study 1. That is, under no load, regardless of testing instructions, participants performed at the same rate. However, if participants were given the high-stakes testing instruction (Study 2), performance was lower under load conditions compared to Study 1, in which there were no additional testing instructions.

Subjective arousal (i.e., SAM score) was not significantly different across studies, except for under an arousing load. Under an arousing load, subjective arousal was found to be significantly higher in Study 2 compared to Study 1. This finding of higher subjective arousal under arousing load under high-stakes testing instructions paired with significantly lower accuracy under load conditions in Study 2, may suggest that subjective arousal did play a factor in the reduced accuracy rate under an arousing load (though not significantly). Additionally, given that subjective arousal was higher in the arousing word condition (in Study 2) versus the neutral word condition suggests that the increase in arousal may be due to the nature of the arousing words, which further supports the overall finding of reduced accuracy in the arousing load condition compared to neutral load or no load conditions. However, the fact that subjective arousal was higher in the arousing condition in Study 2 versus Study 1 suggests that the high-stakes testing instructions may also have played a role in the finding of increased subjective arousal, although the comparison of SAM scores between blocks did not reveal a significant increase in subjective arousal after the high-stakes testing instructions were given. Given those observations, it is plausible that both the arousing nature of the arousing words and the arousal from the high-stakes testing instructions had an additive effect to reduce performance rates in the arousing load condition in Study 2. However, analyses did not show a significant effect of

subjective arousal on performance. This may be due to the subjective nature of self-report measures, and is further discussed below.

5 GENERAL DISCUSSION

The purpose for the current study was to test whether the type of cognitive load (emotional or neutral) differentially influenced reading comprehension, indexed as test performance, as a function of TA, subjective arousal, and WMC. In Study 1, I investigated the potential interaction of these variables in a clean, lab-based manner, which was meant to provide a foundation for the subsequent, more ecologically valid study (Study 2). That is, performance in a no stress testing situation (Study 1) was meant as a comparison for performance under stressful testing conditions (Study 2). Although, Study 1 suffered from low power, an effect of load type was still consistently evident which speaks to the robustness of this effect. This effect of load type was replicated in Study 2.

Although no significant effects of WMC, TA or subjective arousal were found in either study, there was a consistent finding across analyses in Studies 1 and 2 of reduced accuracy under load conditions versus reading alone, as well as reduced accuracy under an arousing load beyond decrements seen under a general cognitive load. A common thread in the theories discussed above (e.g., Baddeley and Engle WM models, ACT, PET, etc.) is the prominent role of attention, and supplies the most probable explanation for the current findings.

The current studies revealed a consistent effect of load type. That is, how well participants performed on the reading comprehension task depended on the content of the load such that performance was worse when under load compared to reading alone, and performance was decreased further when the content of the load was arousing rather than neutral. This finding can be explained by differential attention allocation under the different load types. Previous work has consistently shown that WMC predicts reading comprehension (e.g., Daneman & Carpenter, 1980; Engle, et al., 1992; McVay & Kane, 2012). It has also been shown

that not only verbal WM tasks (e.g., reading span) predict reading comprehension, but also non-verbal measures of WMC such as the OPSPAN, which contains numbers (Conway & Engle, 1996; Daneman & Merikle, 1996; Kane et al., 2004). This would suggest that a domain-general factor, such as attention, is predicting reading comprehension (Kane, Bleckley, Conway, & Engle, 2001; Kane, Hambrick, Tuholski, Wilhelm, Payne, & Engle, 2004), which makes sense as attention is important for reading because it is a controlled process (versus an automatic process). Work by Engle and Kane (2004) confirmed that attention is an underlying factor in both WMC and reading comprehension. Similarly, previous research has shown that WMC also predicts standardized test scores such as the SAT (e.g., Engle, et al., 1999; Turner & Engle, 1989). Engle and colleagues (1999) posited this relationship is due to differences in the ability to control attention, explaining that the inability to maintain focus of attention or inhibit irrelevant thoughts during a test is likely to cause decrements in test performance. Work by Unsworth and colleagues confirmed SAT scores are, in large part, due to differences in the ability to control attention (Unsworth, McMillan, Brewer, & Spillers, 2012).

Working memory capacity has also been shown to be important for processing emotional information (e.g., Redick, Heitz, & Engle, 2007; Unsworth, Heitz, & Engle, 2005). For example, several studies have found that threat-related words are more likely to capture attention than neutral or positive words (MacLeod, et al., 1986; McNally, et al., 1990; Segal, et al., 1995; Williams et al., 1996), and that individuals have difficulty disengaging from emotional or threatening stimuli (Fox, et al., 2001). Both Studies 1 and 2 of the current project support these findings. The consistent finding of reduced performance under arousing cognitive load beyond decrements seen under general cognitive load can be explained by a change in attention control such that under dual-task performance (i.e., load conditions), top down control gave way to more

bottom up processing, letting the load and irrelevant thoughts/distracters interfere with processing, and this effect was more pronounced under an emotionally arousing load due to the emotionally arousing nature of the words. That is, the emotionally arousing words drew more attention than the neutral words, and both load conditions (neutral and arousing word) required more attention resources compared to reading alone (no load condition). Participants' attention was captured by the emotionally arousing words, and thus, less attention remained for reading comprehension. As a result, performance suffered. This is consistent with Engle's conceptualization of attention as one's ability to maintain focus in addition to the ability to inhibit irrelevant distracters (Engle, 2002; Engle & Kane, 2004; Kane, et al., 2001), as well as Kahneman's (1973) conceptualization that attention is a limited capacity resource that can be allocated flexibly to various tasks.

Additionally, performance under the different load types is also consistent with Miyake and colleagues' proposed functions of executive control (Miyake, Friedman, Emerson, Witzki, Howerter, 2000). Authors proposed that executive control has 3 functions: inhibition, updating WM, and shifting between tasks. In the current study, participants likely had difficulty shifting between tasks (maintaining words and reading) under load conditions compared to reading alone, with even more difficulty shifting between tasks under an arousing load due to difficulty disengaging from the arousing words. Additionally, participants likely experienced difficulty in preventing (i.e., inhibiting) the arousing nature of the arousing words from interfering with the reading, as well as any additional irrelevant thoughts or distractions. This observation is consistent with the idea that as load increases, one's ability to suppress irrelevant information decreases (e.g., Conway, Tuholski, Shisler, & Engle, 1999). In the current study, participants would have had increased load and reduced ability to suppress under load conditions, but

especially under an arousing load as the arousing nature of the words occupied more attention compared to the neutral words. Lastly, due to the load on attention, participants likely had difficulty in updating WM as they read passages.

The executive attention view of WM emphasizes one's ability to inhibit irrelevant information, and this applies not only to external distraction, but also internally generated distraction (Engle, 2002). Internal distractions such as irrelevant thoughts and worries have been shown to act as an additional cognitive load, using up attention resources and decreasing the ability to suppress irrelevant information (Ilkowska & Engle, 2010). For example, Schooler, Reichle, and Halpern (2004) found that participants that had task-irrelevant thoughts during reading were more likely to have lower reading comprehension scores compared to participants that had little or no task-irrelevant thoughts. Similarly, worrisome thoughts have also been shown to take up available attention resources needed for task performance (Ashcraft & Kirk, 2001; Eysenck & Keane, 1990). Worry, as experienced in testing, and irrelevant thoughts involve inner verbal processing, which takes up available attention resources, and which is likely to be especially problematic when one is completing a task that is also verbal in nature (Gray, et al., 2002; Gray, 2001). In the current study, it may be that the arousing and neutral loads are similar to the different types of load individuals may experience in a testing situation. The neutral load may be similar to irrelevant, non-emotional thoughts test takers may experience, such as errands they need to run, something they need to pick up from the grocery store, or where they are going after their test is finished, and would cause divided attention during test performance. Further, the emotionally arousing load may be similar to the worry individuals experience in a testing situation, pulling even more attention resources away from test performance.

In terms of Baddeley's (1974) conception of WM, previous research has not examined the effect of taxing the PL with an emotionally arousing WMC load specifically while reading. However, previous work has shown that performance may suffer when two concurrent tasks are both verbal (e.g., Derakshan & Eysenck, 1998; MacLeod & Donnellan, 1993), and indeed the current studies support that work as additional verbal load during reading was associated with decreased performance compared to reading alone. However, when the verbal load was arousing, performance decreased further. The finding of differential performance based on the content of the verbal load is informative as to the relationship between the PL and the central executive. If we assume an attention-based explanation for differential performance according to load type, this finding suggests that the involvement of the PL and central executive in processing verbal information cannot be completely divorced. That is, it is not possible to tax the PL without using some amount of attention to do so. If the arousing word load consumed more attention resources compared to the neutral word load, then some amount of attention is needed to maintain verbal information in the PL. That is, you cannot purely tax the PL without involving the central executive (i.e., attention). If you put a load on the PL, you may also be loading the central executive, depending on the task. So, even discussing the findings in terms of a WM model other than the executive attention view of WM, it is clear that attention is the key factor in the differential effects of load type seen in the current study.

Although, an effect of load type was demonstrated across studies, effects of TA, WMC, and subjective arousal on reading comprehension were not apparent. Previous work provides possible explanations for this lack of findings. First, trait anxiety was not shown to impact reading comprehension performance. However, previous research has demonstrated that TA may not affect the effectiveness (e.g., accuracy) of performance, but rather the efficiency of the

performance (amount of resources used considering level of effectiveness; e.g., Berggren & Derakshan, 2012; Derakshan, Ansari, Hansard, Shoker, & Eysenck, 2009; Eysenck, et al., 2007). For example, in a self-paced reading comprehension task, Calvo and colleagues (1994) found that neither concurrent speech nor articulatory suppression differentially affected comprehension performance as a function of anxiety level (effectiveness), although both of these concurrent tasks significantly reduced reading speed and increased the use of reading regressions and articulatory rehearsal for high anxious participants (efficiency). Conversely, when reading time was determined by the experimenter (i.e., reading was not self-paced), Calvo and Eysenck (1996) found that the use of the PL was more important during reading comprehension to persons with high anxiety, as their comprehension significantly decreased when either component of the PL was disrupted as compared to persons with low anxiety. This would suggest that use of the PL is a resource that high anxiety persons use to maintain performance levels. So, it is likely that the lack of an effect of TA in the current study was due to the reading passages not being timed. According to the reading strategies discussed by Calvo and Eysenck, HTA participants in the current study were not likely to have used articulatory rehearsal as a strategy because they were engaging in articulatory suppression, so HTA participants likely used slowed reading speed and increased regressions to maintain performance. That is, as participants were given unlimited time to read, HTA persons could slow their reading and use as many regressions as necessary to complete the task. Although slowed reading pace and additional reading regressions were not efficient (i.e., more resources used given performance rate), HTA persons were able to perform at the same rate as LTA persons. If an effect of TA had been observed, this would support previous work showing that standardized testing does not

accurately reflect ability across groups, and that differential test performance does not pertain just to those that are test anxious, but yet to another group of people, individuals that are HTA.

Second, the lack of finding of an effect of WMC may also be explained by strategy use in LWMC persons. Calvo and colleagues (1994) examined the use of reading strategies (reading was self-paced) across levels of test anxiety and WMC under different types of interference (concurrent speech, articulatory suppression, and reading alone). Similar to the current study, they also failed to find an effect of WMC despite the fact that LWMC persons routinely perform more poorly on reading tasks because they cannot keep as much previously read information in mind to integrate with later sentences and paragraphs. Calvo and colleagues found that adding a secondary task (articulatory suppression, concurrent speech) served as an additional load on WM resources only when regressive eye fixations could not be used. If regressions could be used, LWMC persons performed similarly to HWMC persons. So, as reading passages in the current study were not timed, it is likely that LWMC subjects used reading strategies such as slowed reading speed and increased reading regressions to help them maintain performance comparable to HWMC persons.

Another possibility for the lack of finding of an effect of WMC is that the present reading comprehension task may have been too easy. Just, Carpenter, and Woolley (1982) observed that differences in reading comprehension between HWMC and LWMC persons did not emerge if the task was too easy, with larger differences occurring between HWMC and LWMC persons as the task increased in difficulty. This connects back to the idea that performance is only affected when WMC has been exceeded. If the task is too easy (i.e., demands little or no WMC), then capacity will not be exceeded, and differences in comprehension will not be found between HWMC and LWMC persons. Future research could address this by increasing the length of the

passages as longer passages are more demanding of attention resources. However, this is not likely as the passages and questions were taken from old SAT exams, which should prove sufficiently challenging for this sample. Additionally, a full range of scores were observed in both Studies 1 and 2. Conversely, one might argue that the task was too difficult as evidenced by the low scores in Study 1 and 2. However, this is likely an issue of motivation, rather than task difficulty because SAT questions should be at an appropriate difficulty level for the sample, and given the fact that in both studies, under no load, accuracy was around 70%, which in academic settings, is usually deemed “average”. An average performance, given that participants’ performance had no adverse consequences for poor performance, was not unreasonable to expect. Additionally, it is likely that effects of TA, WMC, and subjective arousal would emerge if participants were in a real testing situation as a real testing situation would likely be much more stressful compared to a laboratory setting, given that individual differences in TA and WMC most frequently appear under stress (e.g., Berggren, & Derakshan, 2013; Calvo, et al., 1994; Eysenck, et al., 2007; Ilkowska & Engle, 2010).

Lastly, subjective arousal did not have an effect on reading comprehension performance. This may be due to the subjective nature of self-report measures. For example, some groups (e.g., women, high anxious) report higher subjective arousal or task difficulty although their performance levels are on par with their counterparts (e.g., men, low anxious) (Berggren & Derakshan, 2013). Also, it is possible that participants (or enough participants) were not aroused enough during lab-based testing. That is, there were no adverse consequences associated with failure on the reading comprehension test and so, emotional arousal was not as high as in a real high-stakes testing situation, and was not at a level high enough to adversely affect performance. Instead, in both Studies 1 and 2, arousal increased systematically over the duration of the study,

and may have influenced the calculation of whether participants were subjectively aroused or not. However, this is less of a concern in Study 1 as several analyses were conducted on the first block of data only. Study 1 also suffered from low power, which may likely also have contributed to the lack of finding of subjective arousal. However, no significant effect of subjective arousal was observed in Study 2, which was sufficiently powered, and it is reasonable to expect that increased arousal would be more likely to appear under stress conditions, so Study 2 was the more likely of the two studies to show an effect of subjective arousal. In Study 2, even though most comparisons revealed an increase in arousal across time, no significant increase was observed after the high-stakes testing instructions had been given. However, it is likely that the high-stakes testing instructions provided some motivation (though not arousal) as evidenced by the lack of cognitive fatigue effects in Study 2. I would expect to find a significant increase in reading comprehension performance if participants were actually in a real high-stakes testing situation. Future studies may add physiological measures such as heart rate, galvanic skin response, or cortisol measurements to supplement self-report arousal measures.

To summarize, despite cognitive fatigue effects and low power, Study 1 revealed an effect of load type such that performance was worse under dual-task performance. Additionally, performance under load conditions depended on the content of the load, such that reading comprehension performance was lower when participants were under an arousing load compared to a neutral load. This effect demonstrated that the load manipulation was, in fact, working, and provided a platform from which to conduct Study 2. Consistent with Study 1, an effect of load type was also found in Study 2, with performance being worse under load conditions, but worse still in the arousing load condition compared to the neutral load condition. The lack of findings of effects of TA and WMC are likely due to the reading passages not being timed; whereas, the

lack of finding of an effect of subjective arousal was likely a result of participants not being aroused enough (like in a real high-stakes test) to cause decrements in performance.

The results from the current study suggest that an emotional load, such as experienced in high-stakes testing, may not unduly single out persons based on level of WMC, TA, or subjective arousal, causing lower scores than motivation and ability would predict. That is, the emotional nature of the testing environment may not differentially affect individuals differing in levels of WMC, TA, or subjective arousal, causing an increased likelihood of “choking” under pressure. This is certainly encouraging as this would mean that testing is fair across individuals, and scores may reflect actual ability. However, more research is needed, addressing the shortcomings of the current study, to verify this possibility.

5.1 Future Directions

To determine effects of WMC and TA on performance, future studies should use reading passages are *actually* timed rather than just *advising* participants the passages are timed. Under a timed condition, LWMC and HTA persons will have limited strategy use (i.e., they will not be able to use slowed reading or increased reading regressions), and therefore effectiveness may suffer as a result rather than just processing efficiency, and differences in performance between LWMC and HWMC persons, as well as differences between LTA and HTA persons, may emerge. However, if future studies find no effect of WMC, it might be concluded that HWMC is not protective against the effects of an emotional load.

To increase face validity, future studies may use words associated with good performance (e.g., pass) versus poor performance (e.g., fail) on tests, with a similar design as the current study. Such words may better simulate the thoughts that students may experience in a testing situation such as confidence and viewing the test as a challenge versus anxiety and fear of

failure. Additionally, future studies may give participants something from their own lives to worry about in order to simulate worry in a testing situation. Asking participants to worry about something from their own lives has the advantage of the worry being personally relevant to the participant, perhaps resulting in a better simulation of the worry experienced in testing.

Lastly, in a series of studies, Logel, Iserman, Davies, Quinn, and Spencer (2003) manipulated the thought content of women under stereotype threat (math) by having some participants replace any negative thoughts about the math test with thoughts regarding an important social identity (Study 4) or irrelevant thoughts (a red Volkswagon; Study 5). The results showed that participants under stereotype threat that replaced thoughts about stereotype with positive or irrelevant thoughts performed better than women that did not receive the thought replacement instructions, and also performed as well as men. Future studies may investigate the utility of this thought replacement technique during reading comprehension with high-stakes testing procedures as in the current study.

5.2 Conclusion

In this study, I examined WMC limitations in test performance by using arousing and non-arousing loads to simulate the types of load or distraction experienced in a testing situation. In both studies 1 and 2, reading comprehension performance depended on the content of the load. Under divided attention (load conditions), performance decreased compared to reading alone; however, more attention resources were further consumed by the nature of the arousing words in the arousing load.

Additionally, Study 1 showed decreasing accuracy as a function of experiment length; however, under induced state anxiety (Study 2), participants were able to muster cognitive resources and control attention to improve performance (i.e., not succumb to cognitive fatigue

effects as in Study 1). Overall, differences in the pattern of performance across studies underscore the importance of task characteristics and contextual information in test performance.

Lastly, rehearsing information with differing affective values may be an effective method with which to simulate irrelevant and worrisome thoughts like those experienced in a testing environment. However, replication is needed to verify this possibility.

REFERENCES

- Ashcraft, M. H., & Kirk, E. P. (2001). The relationships among working memory, math anxiety, and performance. *Journal of Experimental Psychology: General*, 130, 224 – 237.
- Baddeley, A. D. (2003). Working memory and language; an overview. *Journal of Communication Disorders*, 36, 189 – 208.
- Baddeley, A. D. (2007). *Working memory, thought and action*. Oxford, UK: Oxford University Press.
- Baddeley, A., Banse, R., Huang, Y., Page, M. (2012). Working memory and emotion: Detecting the hedonic detector. *Journal of Cognitive Psychology*, 24 (1), 6 – 16.
- Baddeley, A. D., Eldridge, M., & Lewis, V. (1981). The role of subvocalisation in reading. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 33 (4), 439 – 454.
- Baddeley, A. D., Gathercole, S., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review*, 105 (1), 158 – 173.
- Baddeley, A. D. & Hitch, G. J. (1974). Working memory. In G. Bower (Ed.). *The psychology of learning and motivation* (Vol. 8, pp. 47 – 90). New York: Academic Press.
- Baddeley, A. D., Logie R., Nimmo-Smith, I., & Brereton, N. (1985). Components of fluent reading. *Journal of Memory and Language*, 24, 119 – 131.
- Baddeley, A. D., Thomson, N., Buchanan, M. (1975b). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14, 575 – 589.
- Bayer, M., Sommer, W., & Schacht, A. (2011). Emotional words impact the mind but not the body: Evidence from pupillary responses. *Psychophysiology*, 48, 1553 – 1561.
- Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psy-*

- chological Science*, 17 (5), 339 – 343.
- Beilock, S. L. & Carr, T. H. (2005). When high-powered people fail: Working memory and ‘choking under pressure’ in math. *Psychological Science*, 16, 101 – 105.
- Beilock, S. L. & DeCaro, M. S. (2007). From poor performance to success under stress: Working memory, strategy selection, and mathematical problem solving under pressure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, (6), 983 – 998.
- Beilock, S. L., Kulp, C. A., Holt, L. E., & Carr, T. H. (2004). More on the fragility of performance: Choking under pressure in mathematical problem solving. *Journal of Experimental Psychology: General*, 133, 584 – 600.
- Beilock, S. L., Rydell, R. J., & McConnell, A. R. (2007). Stereotype threat and working memory: Mechanisms, alleviation, and spillover. *Journal of Experimental Psychology: General*, 136, 256 – 276.
- Berggren, N. & Derakshan, N. (2012). Attentional control deficits in trait anxiety: Why you see them and why you don’t. *Biological Psychology*, 92 (3), 440 – 446.
- Bradley, M. M. & Lang, P. J. (1999). *Affective norms for English words (ANEW): Instruction manual and affective ratings*. Technical Report C-1, The Center for Research in Psychophysiology, University of Florida.
- Broadway, J. M. & Engle, R. W. (2010). Validating running memory span: Measurement of working memory capacity and links with fluid intelligence. *Behavior Research Methods*, 42, 563 – 570.
- Brown, J. I., Vick-Fishco, V., Hanna, G. S. (1993). *Nelson-Denny Reading Test*. Rolling Meadows, IL: Riverside.
- Calvo, M. G. & Eysenck, M. W. (1996). Phonological working memory and reading in test

- anxiety. *Memory*, 4 (3), 289 – 305.
- Calvo, M. G., Eysenck, M. W., Estevez, A. (1994). Ego-threat interpretive bias in test anxiety: On-line inferences. *Cognition and Emotion*, 8 (2), 127 – 146.
- Calvo, M. G., Eysenck, M. W., Ramos, P. M., & Jimenez, A. (1994). Compensatory reading strategies in test anxiety. *Anxiety, Stress, and Coping*, 7, 99 – 116.
- Calvo, M. G., Ramos, P. M., Estevez, A. (1992). Test anxiety and comprehension efficiency: the role of prior knowledge and working memory deficits. *Anxiety, Stress, and Coping*, 5, 125 – 138.
- Caplan, D. & Waters, G. (1995). Aphasic disturbances of syntactic comprehension and working memory capacity. *Cognitive Neuropsychology*, 12, 637 – 649.
- Cassady, J. C. & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. *Contemporary Educational Psychology*, 27, 270 – 295.
- Chen, I. & Chang, C. (2009). Cognitive load theory: an empirical study of anxiety and task performance in language learning. *Electronic Journal of Research in Educational Psychology*, 7 (2), 729 – 746.
- Christopher, M. E., Miyake, A., Keenan, J. M., Pennington, B., DeFries, J. C., Wadsworth, S. J., Willcutt, E., & Olson, R. K. (2012). Predicting word reading and comprehension with executive function and speed measures across development: a latent variable analysis. *Journal of Experimental Psychology: General*, 141, (3), 470 – 488.
- Cohen, J. D., Dunbar, K., & McClelland, J. (1990). On the control of automatic processes: a parallel distributed processing account of the Stroop effect. *Psychological Review*, 97, 332 – 361.

- Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term-memory recall of repeated and novel word lists. *Memory and Cognition*, 21 (4), 539 – 545.
- Coltheart, V., Avons, S. E., Trollope, J. (1990). Articulatory suppression and phonological codes in reading for meaning. *The Quarterly Journal of Experimental Psychology Section A: Human Experimental Psychology*, 42 (2), 375 – 399.
- Conway, A. R. A. & Engle, R. W. (1996). Individual differences in WM capacity: More evidence for a general capacity theory. *Memory*, 4, 577 – 590.
- Cowan, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Ed.), *Models of Working Memory* (pp. 62 – 101). Cambridge, UK: Cambridge University Press.
- Daneman, M. & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior*, 19, 450 – 466.
- Daneman, M. & Merikle, P. M. (1996). Working memory and language comprehension: a meta-analysis. *Psychonomic Bulletin & Review*, 3, 422 – 433.
- Darke, S. (1988a). Effects of anxiety on inferential reasoning task performance. *Journal of Personality and Social Psychology*, 55, 499 – 505.
- Deffenbacher, J. L. (1980). Worry and emotionality in test anxiety. In I. G. Sarason, (Ed.), *Test anxiety: Theory, research, and applications* (pp. 111 – 124). Hillsdale, NJ: Erlbaum.
- Derakshan, N., Ansari, T. L., Hansard, M., Shoker, L., Eysenck, M.W. (2009). Anxiety, inhibition, efficiency, and effectiveness. *Experimental Psychology*, 56 (1), 48 – 55.
- Derakshan, N. & Eysenck, M. W. (1998). Working memory capacity in high trait-anxious and repressor groups. *Cognition and Emotion*, 12, 697 – 713.

- Derakshan, N. & Eysenck, M. W. (2010). Introduction to the special issue: Emotional states, attention, and working memory. *Cognition and Emotion*, 24 (2), 189 – 199.
- Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review*, 66 (3), 183 – 199.
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11 (1), 19 – 23.
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: a test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18, 972 – 992.
- Engle, R. W. & Kane, M. J. (2004). Executive attention, working memory capacity and two-factor theory of cognitive control. In B. Ross (Ed.), *The Psychology of Learning and Motivation* (pp. 145 – 199). New York: Elsevier.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: a latent variable approach. *Journal of Experimental Psychology: General*, 128, 309 – 331.
- Eysenck, M. W. (1979). Anxiety, learning and memory: a reconceptualization. *Journal of Research in Personality*, 13, 363 – 385.
- Eysenck, M. W. & Byrne, A. (1994). Implicit memory bias, explicit memory bias, and anxiety. *Cognition and Emotion*, 8 (5), 415 – 431.
- Eysenck, M. W. & Calvo, M. G. (1992). Anxiety and performance: the processing efficiency theory. *Cognition and Emotion*, 6, 409 – 434.
- Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, 7 (2), 336 – 353.

- Eysenck, M. W. & Keane, M. T. (1990). *Cognitive psychology: a student's handbook*. Hillsdale, NJ: Erlbaum.
- Eysenck, M. W., Payne, S., & Derakshan, N. (2005). Trait anxiety, visuospatial processing, and working memory. *Cognition and Emotion*, 19, 1214 – 1228.
- Fartoukh, M., Chanquoy, L., & Piolat, A. (2012). Effects of emotion of writing processes in children. *Written Communication*, 29 (4), 391 – 411.
- Fox, E., Russo, R., Bowles, R., & Dutton, K. (2001). Do threatening stimuli draw or hold visual attention in subclinical anxiety? *Journal of Experimental Psychology: General*, 130, 681 – 700.
- Gimmig, D., Huguet, P., Caverni, J. P., & Cury, F. (2006). Choking under pressure and working memory capacity: When performance pressure reduces fluid intelligence. *Psychonomic Bulletin & Review*, 13, 1005 – 1010.
- Gotoh, F. (2008). Influence of affective valence on working memory processes. *International Journal of Psychology*, 43 (1), 59 – 71.
- Gray, J. R. (2001). Emotional modulation of cognitive control: Approach-withdrawal states double-dissociate spatial from verbal two-back task performance. *Journal of Experimental Psychology: General*, 130, 436 – 452.
- Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences, USA*, 99, 4115 – 4120.
- Gruber, G. R. (2011). *Gruber's Complete SAT Critical Reading Workbook*. Sourcebooks, Inc.: Naperville, IL.
- Huang, Y. M., Baddeley, A., & Young, A. W. (2008). Attentional capture by emotional stimuli is

- modulated by semantic processing. *Journal of Experimental Psychology: Human Perception and Performance*, 34 (2), 328 – 339.
- Ilkowska, M. & Engle, R. W. (2010). Trait and state differences in working memory capacity. In A. Gruszka, G. Matthews, & B. Szymura (Eds.), *Handbook of individual differences in cognition: Attention, memory and executive control*. New York: Springer.
- Jimenez-Ortega, L. Martin-Loeches, M., Casado, P., Sel, A., Fondevila, S., Herreros de Tejada, A., Schacht, A., & Sommer, W. (2012). *How the emotional content of discourse affects language comprehension*. PLoS ONE, 7 (3): e33718. doi: 10.1371/journal.pone.0033718
- Just, M. A. & Carpenter, P. A. (1992). A capacity theory of comprehension: Individual differences in working memory. *Psychological Review*, 99 (1), 122 – 149.
- Just, M. A., Carpenter, P. A., & Woolley, J. D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, 228 – 238.
- Kahneman, D. (1973). *Attention and effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kane, M. J., Bleckley, M. K., Conway, A. R. A., & Engle, R. W. (2001). A controlled-attention view of WM capacity. *Journal of Experimental Psychology: General*, 130, 169 – 183.
- Kane, M. J. & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 26 (2), 336 -358.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The domain generality of working-memory capacity: a latent-variable approach to verbal and spatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133, 189 – 217.
- King, J. & Just, M. A. (1991). Individual differences in syntactic processing: the role of working

- memory. *Journal of Memory and Language*, 30, 580 – 602.
- Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H., Johnson, and T. A. Williams (Eds.), *Technology in Mental Health Care: Delivery Systems* (pp. 119 – 137). Norwood, N. J.: Ablex.
- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, 19 (3), 143 – 148.
- Lavie, N., Hirst, A., de Fockert, J. W., Viding, E. (2004). Load theory of selective attention and cognitive control. *Journal of Experimental Psychology: General*, 133 (3), 339 – 354.
- Levens, S. M. & Phelps, E. A. (2008). Emotion processing effects on interference resolution in working memory. *Emotion*, 8 (2), 267 – 280.
- Liebert, R. M. & Morris, L. W. (1967). Cognitive and emotional components of test anxiety: a distinction and some initial data. *Psychological Reports*, 20, 975 – 978.
- Lindstrom, B. R. & Bohlin, G. (2011). Emotion processing facilitates working memory performance. *Cognition and Emotion*, 25 (7), 1196 – 1204.
- Logel, C., Iserman, E. C., Davies, P. G., Quinn, D. M., & Spencer, S. J. (2009). The perils of double consciousness: the role of thought suppression in stereotype threat. *Journal of Experimental Social Psychology*, 45, 299- 312.
- MacLeod, C. & Donnellan, A. M. (1993). Individual differences in anxiety and the restriction of working memory capacity. *Personality Individual Differences*, 15 (2), 163 – 173.
- MacLeod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology*, 95, 15 – 20.
- Mangels, J. A., Good, C., Whiteman, R. C., Maniscalco, B., Dweck, C. S. (2012). Emotion blocks the path to learning under stereotype threat. *Social, Cognitive, & Affective*

- Neuroscience*, 7, 230 – 241.
- Mattarella-Micke, A., Mateo, J., Kozak, M. N., Foster, K., Beilock, S. L. (2011). Choke or thrive? The relation between salivary cortisol and math performance depends on individual differences in working memory and math anxiety. *Emotion*, 11 (4), 1000 – 1005.
- Matthews, G. & Campbell, S. E. (2010). Dynamic relationships between stress states and working memory. *Cognition and Emotion*, 23 (2), 357 – 373.
- McNally, R. J., Riemann, B. C., & Kim, E. (1990). Selective processing of threat cues in panic disorder. *Behavior Research and Therapy*, 28, 407 – 412.
- McVay, J. C. & Kane, M. J. (2012). Why does working memory capacity predict variation in reading comprehension? On the influence of mind wandering and executive attention. *Journal of Experimental Psychology: General*, 141 (2), 302 – 320.
- Mueller, J. H. & Overcast, T. D. (1976). Free recall as a function of test anxiety, concreteness, and instructions. *Bulletin of the Psychonomic Society*, 8, 194 – 196.
- Naveh-Benjamin, M., Maddox, G. B., Jones, P., Old, S., & Kilb, A. (2012). The effects of emotional arousal and gender on the associative memory deficits of older adults. *Memory and Cognition*, 40, 551 – 566.
- Rai, M. K., Loschky, L. C., Harris, R. J., Peck, N. R., Cook, L. G. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning*, 61 (1), 187 – 218.
- Rapee, R. M. (1993). The utilization of working memory by worry. *Behaviour Research and Therapy*, 31, 617 – 620.

- Redick, T. S., Heitz, R. P., & Engle, R. W. (2007). Working memory capacity and inhibition: Cognitive and social consequences. In D. S. Gorfein & C. M. MacLeod (Eds.), *Inhibition in Cognition* (pp. 125 – 142). Washington, D.C.: American Psychological Association.
- Rosen, V. M. & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, 126 (3), 211 – 227.
- Schackman, A. J., Sarinopoulos, I., Maxwell, J. S., Pizzagalli, D. A., Lavric, A., Davidson, R. J. (2006). Anxiety selectively disrupts visuospatial working memory. *Emotion*, 6, 40 – 61.
- Scherbaum, C. A. & Goldstein, H. W. (2008). Examining the relationship between race-based differential item functioning and item difficulty. *Educational and Psychological Measurement*, 68 (4), 537 – 553.
- Schmader, T. & Johns, M. (2003). Converging evidence that stereotype threat reduces working memory capacity. *Journal of Personality and Social Psychology*, 85 (3), 440 – 452.
- Schmader, T., Johns, M., & Forbes, C. (2008). An integrated process model of stereotype threat effects on performance. *Psychological Review*, 115 (2), 336 – 356.
- Schooler, J. W., Reichle, E. D., & Halpern, D. V. (2004). Zoning out while reading: Evidence for dissociations between experience and metaconsciousness. In D. Levin (Ed.), *Thinking and seeing: Visual metacognition in adults and children* (pp. 203 – 226). Cambridge, MA: MIT Press.
- Schwarzer, R. & Jerusalem, M. (1992). Advances in anxiety theory: a cognitive process approach. In K. A. Hagtvet & T. B. Johnsen (Eds.), *Advances in test anxiety research* (Vol. 7, pp. 2 – 31). Lisse, The Netherlands: Swets & Zeitlinger.

- Segal, Z. V., Gemar, M., Truchon, C., Guirguis, M., & Horowitz, L. M. (1995). A priming methodology for studying self-representation in major depressive disorder. *Journal of Abnormal Psychology, 104*, 205 – 213.
- Sliwinski, M. J., Smyth, J. M., Hofer, S. M., & Stawski, R. S. (2006). Intraindividual coupling of daily stress and cognition. *Psychology and Aging, 21*, 545 – 557.
- Spielberger, C. D. (1966). Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior*. New York: Academic Press.
- Spielberger, C. D. Gorsuch, R. L., Lushene, R., Vagg, P. R., & Jacobs, G. A. (1983). *Manual for the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science: A Multidisciplinary Journal, 12* (2), 257 – 285.
- Turner, M. L. & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language, 28*, 127 – 154.
- Unsworth, N., Heitz, R. P., & Engle, R. W. (2005). Working memory capacity in hot and cold cognition. In R. W. Engle, G. Sedek, U. von Hecker, & D. N. McIntosh (Eds.), *Cognitive Limitations in Aging and Psychopathology* (pp. 19 – 43). New York: Cambridge University Press.
- Unsworth, N., McMillan, B. D., Brewer, G. A., Spillers, G. J. (2012). Everyday attention failures: an individual differences investigation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38* (6), 1765 – 1772.
- Walpole, M., Burton, N. W., Kanyi, K., Jackenthal, A. (2002). Selecting successful graduate students: In-depth interviews with GRE users. *GRE Research*. GRE Board Research Report 99-11R. Educational Testing Service: Princeton, NJ.

- Waters, G. S. & Caplan, D. (1996). The measurement of verbal working memory capacity and its relation to reading comprehension. *The Quarterly Journal of Experimental Psychology*, 49 (1), 51 – 79.
- Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin*, 120, 3 – 24.
- Wilson, M. S. (1988). The MRC Psycholinguistic Database: Machine Readable Dictionary, Version 2. *Behavioural Research Methods, Instruments and Computers*, 20 (1), 6-11.
- Wine, J. (1980). Cognitive-attentional theory of test anxiety. In Sarason, I. G. (Ed.), *Test anxiety: Theory, research and applications* (pp. 349 – 385). Hillsdale, NJ: Erlbaum.
- Yerkes, R. M. & Dodson, J. D. (1908). The relation of strength of stimulation to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology*, 18, 459 – 482.
- Yiend, J. & Mathews, A. (2001). Anxiety and attention to threatening pictures. *The Quarterly Journal of Experimental Psychology*, 54A, 665 – 681.

APPENDICES

APPENDIX A: TABLES AND FIGURES

Table 1. Mean SAM Scores by Block Order.

	Block 1 <i>M</i> (<i>SD</i>)	Block 2 <i>M</i> (<i>SD</i>)	Block 3 <i>M</i> (<i>SD</i>)
Order 1:	no load 3.81(2.00)	neutral 3.53 (2.13)	arousing 3.83 (2.18)
Order 2:	neutral 3.72 (2.10)	no load 4.34 (2.51)	arousing 4.83 (2.22)
Order 3:	arousing 3.39 (2.10)	no load 2.64 (1.92)	neutral 3.57 (2.44)
Order 4:	no load 3.30 (1.73)	arousing 3.59 (2.04)	neutral 3.63 (1.78)
Order 5:	neutral 2.91 (2.02)	arousing 3.41 (2.08)	no load 3.47 (2.21)
Order 6:	arousing 2.83 (1.55)	neutral 3.27 (1.69)	no load 3.02 (1.89)

Table 2. Mean SAM Score by Block.

Block		
Block 1	Block 2	Block 3
3.30 (1.93)	3.46 (2.09)	3.68 (2.17)

Note. Values in parentheses are standard deviations.

Table 3. Paired-sample t-test Results for SAM Scores by Block.

Comparison	Mean Difference	Standard Error Mean	<i>t</i>
Block 1 vs. Block 2	-.15	.14	-1.14
Block 1 vs. Block 3	-.38	.15	-2.55*
Block 2 vs. Block 3	-.23	.13	-1.74

Note. * indicates significance at .05 level.

Table 4. Mean Accuracy by Block Order.

	Block 1 <i>M (SD)</i>	Block 2 <i>M (SD)</i>	Block 3 <i>M (SD)</i>
Order 1:	no load .70 (.16)	neutral .53 (.19)	arousing .44 (.18)
Order 2:	neutral .65 (.18)	no load .53 (.15)	arousing .36 (.21)
Order 3:	arousing .62 (.20)	no load .54 (.22)	neutral .42 (.17)
Order 4:	no load .69 (.14)	arousing .53 (.15)	neutral .49 (.19)
Order 5:	neutral .71 (.16)	arousing .58 (.17)	no load .50 (.17)
Order 6:	arousing .65 (.22)	neutral .53 (.19)	no load .43 (.20)

Table 5. Mean Accuracy by Block.

Block		
Block 1	Block 2	Block 3
.67 (.18)	.54 (.18)	.44 (.19)

Note. Values in parentheses are standard deviations.

Table 6. Paired-sample t-test Results for Accuracy by Block.

Comparison	Mean Difference	Standard Error Mean	<i>t</i>
Block 1 vs. Block 2	.13	.01	8.92**
Block 1 vs. Block 3	.23	.02	14.38**
Block 2 vs. Block 3	.09	.01	6.63**

Note. * indicates significance at .05 level. ** indicates significance at .01 level.

Table 7. Marginal Means for Comprehension Accuracy by Load Type.

Load Type		
No Load	Neutral Load	Arousing Load
.70 (.03)	.70 (.02)	.60 (.03)

Note. Values in parentheses are standard error estimates.

Table 8. Marginal Means for Comprehension Accuracy by WMC, Subjective Arousal, and Load Type.

Load Type	WMC	Subjective Arousal	
		Not Aroused	Aroused
No Load	Low	.71 (.04)	.71 (.06)
	High	.68 (.05)	.69 (.06)
Neutral Load	Low	.76 (.05)	.66 (.06)
	High	.70 (.05)	.69 (.05)
Arousing Load	Low	.54 (.05)	.67 (.06)
	High	.72 (.05)	.56 (.09)

Note. Values in parentheses are standard error estimates.

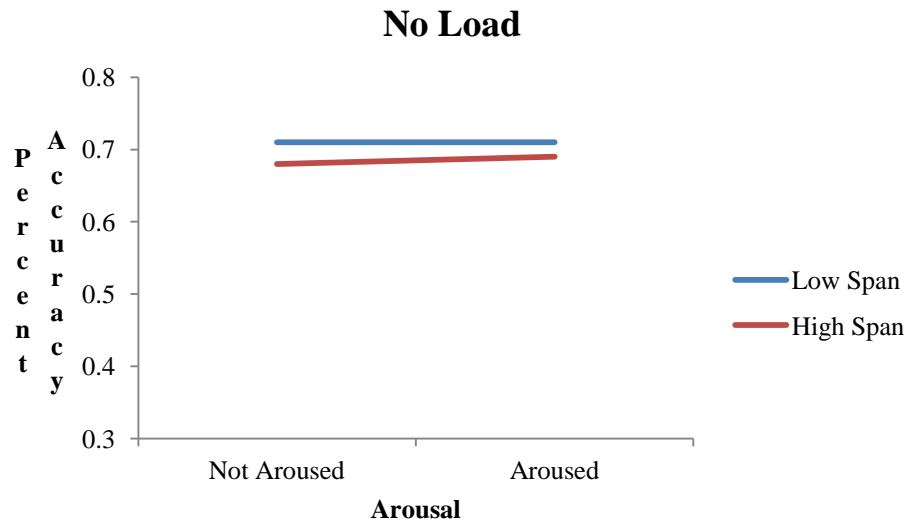


Figure 1. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under No Load.

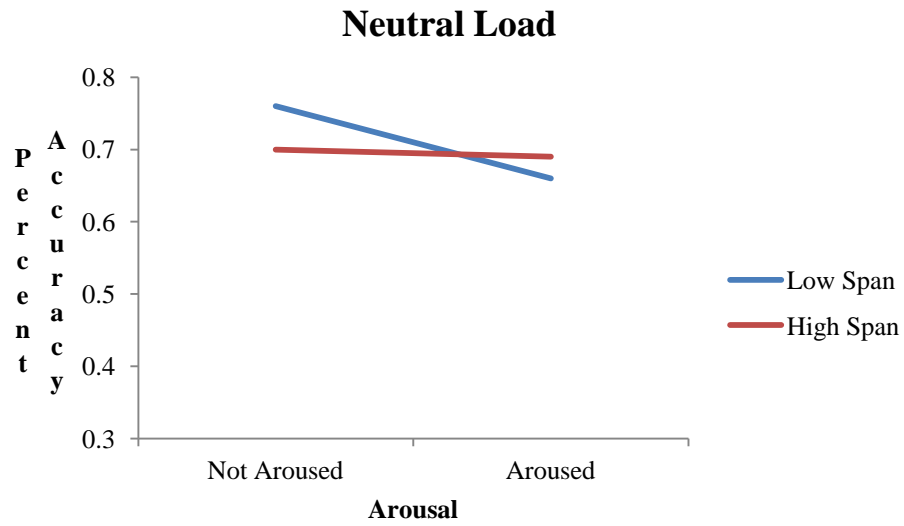


Figure 2. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Neutral Load.

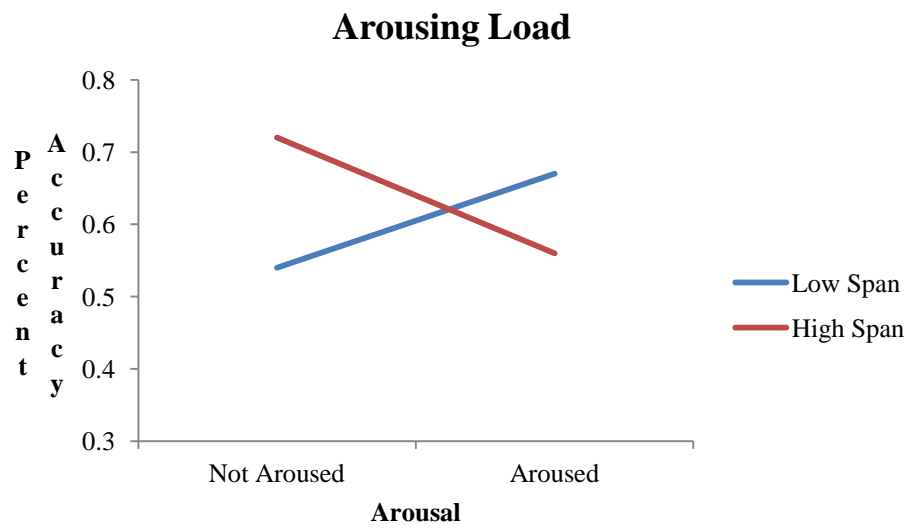


Figure 3. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Arousing Load.

Table 9. Marginal Means for WMC, Subjective Arousal, and Load Type.

Variable	Mean	Standard Error
<u>WMC</u>		
Low	.68	.02
High	.67	.03
<u>Subjective Arousal</u>		
Aroused	.66	.03
Not Aroused	.69	.02
<u>Load Type</u>		
No Load	.70	.03
Neutral Load	.70	.03
Arousing Load	.62	.03

Table 10. Marginal Means for Comprehension Accuracy by Load Type.

Load Type		
No Load	Neutral Load	Arousing Load
.70 (.03)	.72 (.03)	.61 (.03)

Note. Values in parentheses are standard error estimates.

Table 11. Marginal Means for Comprehension Accuracy by WMC, Subjective Arousal, and Load Type.

Load Type	WMC	Subjective Arousal	
		Not Aroused	Aroused
No Load	Low	.54 (.03)	.63 (.04)
	High	.57 (.03)	.54 (.04)
Neutral Load	Low	.57 (.04)	.52 (.05)
	High	.56 (.04)	.61 (.05)
Arousing Load	Low	.53 (.03)	.52 (.04)
	High	.55 (.03)	.49 (.04)

Note. Values in parentheses are standard error estimates.

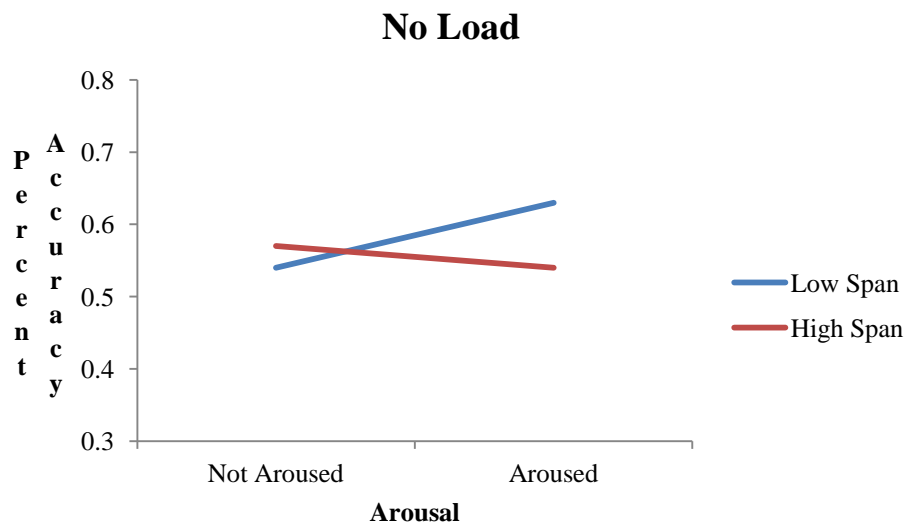


Figure 4. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under No Load.

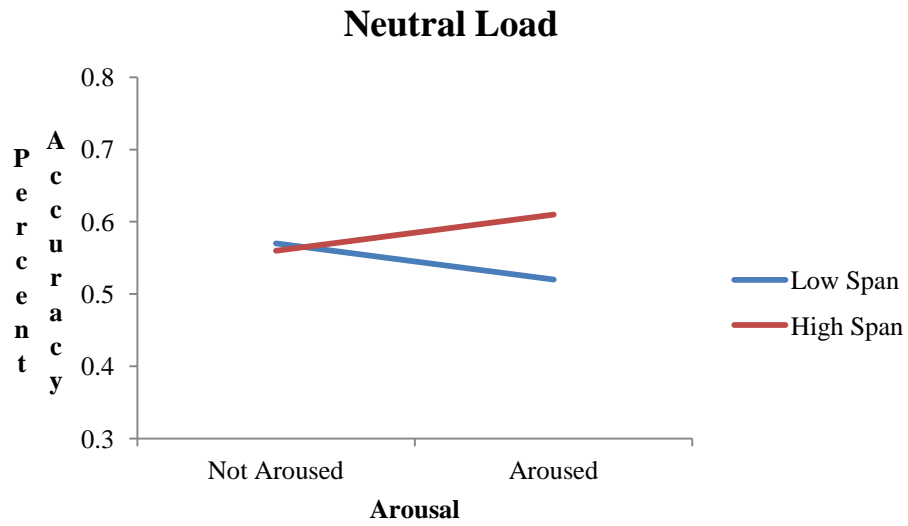


Figure 5. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Neutral Load.

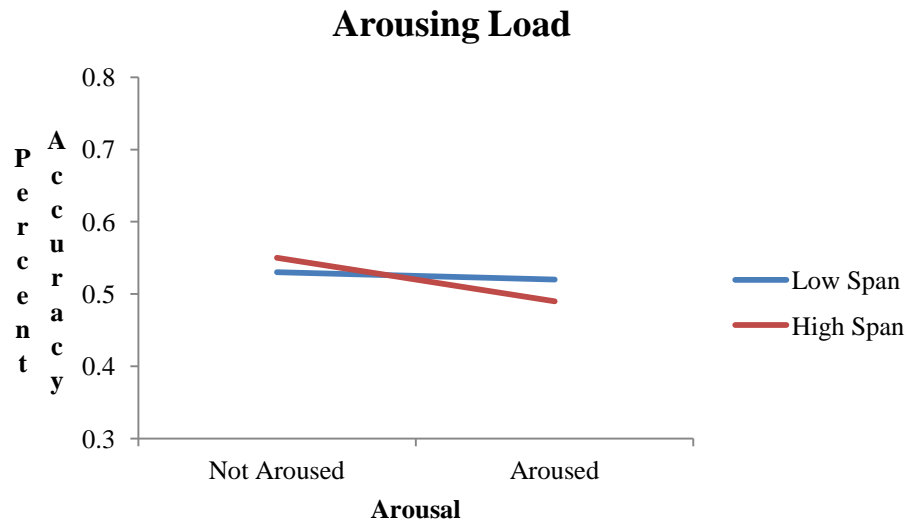


Figure 6. Marginal Means for Comprehension Accuracy by WMC and Subjective Arousal under Arousing Load.

Table 12. Marginal Means for WMC, Subjective Arousal, and Load Type.

Variable	Mean	Standard Error
<u>WMC</u>		
Low	.56	.02
High	.56	.02
<u>Subjective Arousal</u>		
Aroused	.55	.02
Not Aroused	.55	.01
<u>Load Type</u>		
No Load	.57	.01
Neutral Load	.56	.02
Arousing Load	.51	.02

Table 13. Means for Accuracy by Ethnicity.

Ethnicity	Mean Accuracy	Standard Error
Caucasian	.72	.03
African-American	.64	.02
Hispanic	.70	.04
Asian	.64	.03
Other	.70	.03

Table 14. Mean SAM Score by Block Order.

	Baseline <i>M (SD)</i>	Block 1 <i>M (SD)</i>	Block 2 <i>M (SD)</i>	Block 3 <i>M (SD)</i>
Order 1:	2.67 (1.38)	no load 2.96 (1.78)	neutral 3.29 (2.10)	arousing 3.36 (2.23)
Order 2:	3.23 (1.69)	neutral 3.52(1.97)	no load 3.57 (1.87)	arousing 3.75(1.95)
Order 3:	3.37 (1.93)	arousing 3.73(1.79)	no load 3.62 (2.11)	neutral 3.87 (2.07)
Order 4:	3.53 (2.02)	no load 3.47 (1.92)	arousing 3.49 (1.79)	neutral 3.61 (2.12)
Order 5:	3.51 (2.08)	neutral 3.73 (2.12)	arousing 3.71 (2.08)	no load 3.29 (1.97)
Order 6:	4.00 (2.03)	arousing 4.41(1.91)	neutral 4.15 (2.12)	no load 3.87 (2.11)

Table 15. Mean SAM Score by Block.

Block			
Baseline	Block 1	Block 2	Block 3
3.43 (1.92)	3.49 (1.97)	3.71 (2.09)	3.76 (1.96)

Note. Values in parentheses are standard deviations.

Table 16. Paired-sample t-test Results for SAM Score by Block.

Comparison	Mean Difference	Standard Error Mean	<i>t</i>
Baseline vs. Block 1	-.07	-.10	-.67
Baseline vs. Block 2	-.28	.10	-2.73*
Baseline vs. Block 3	-.33	.10	-3.30**
Block 1 vs. Block 2	-.22	.08	-2.63*
Block 1 vs. Block 3	-.27	.08	-3.40**
Block 2 vs. Block 3	-.05	.09	-.55

Note. * indicates significance at .05 level. ** indicates significance at .01 level.

Table 17. Mean Accuracy by Block Order.

	Block 1 <i>M (SD)</i>	Block 2 <i>M (SD)</i>	Block 3 <i>M (SD)</i>
Order 1:	no load .71 (.18)	neutral .51 (.19)	arousing .47 (.19)
Order 2:	neutral .50 (.15)	no load .69 (.18)	arousing .44 (.20)
Order 3:	arousing .49 (.18)	no load .71 (.20)	neutral .51 (.19)
Order 4:	no load .70 (.16)	arousing .48 (.19)	neutral .54 (.15)
Order 5:	neutral .56 (.18)	arousing .44 (.21)	no load .69 (.20)
Order 6:	arousing .46 (.16)	neutral .54 (.16)	no load .68 (.17)

Table 18. Mean Accuracy by Block.

Block		
Block 1	Block 2	Block 3
.57 (.20)	.56 (.21)	.56 (.20)

Note. Values in parentheses are standard deviations.

Table 19. Paired-sample t-test Results for Accuracy by Block.

Comparison	Mean Difference	Standard Error Mean	<i>t</i>
Block 1 vs. Block 2	.01	.01	.40
Block 1 vs. Block 3	.01	.01	.78
Block 2 vs. Block 3	.004	.01	.31

Note. * indicates significance at .05 level. ** indicates significance at .01 level.

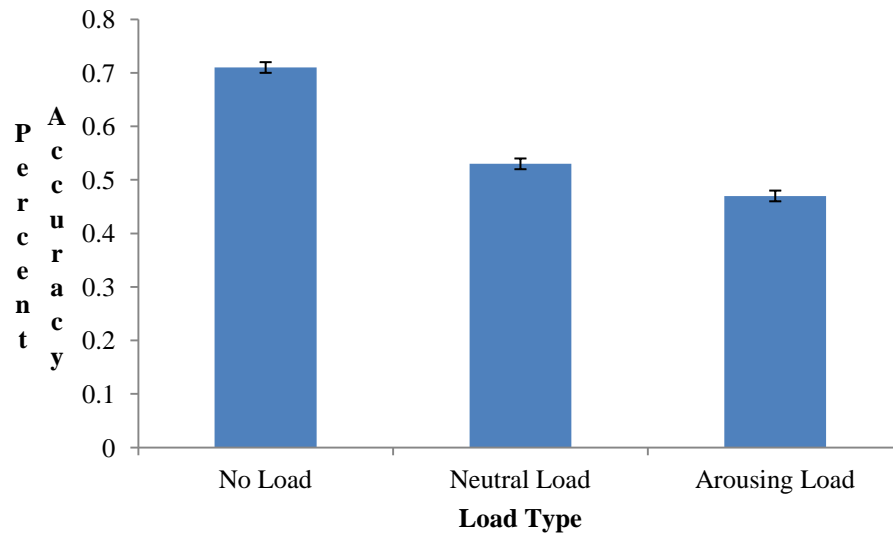


Figure 7. Marginal Means for Comprehension Accuracy by Load Type.

Table 20. Means for Accuracy by Ethnicity.

Ethnicity	Mean Accuracy	Standard Error
Caucasian	.62	.02
African-American	.56	.01
Hispanic	.60	.04
Asian	.50	.03
Other	.57	.03

APPENDIX B: WORD STIMULI

Arousing words	Neutral words
Panic	Statue
Abuse	Engine
Cancer	Guitar
Danger	Ballot
Victim	Extra
Rifle	Rattle
Chaos	Candle
Hatred	Cargo
Poison	Leader
Afraid	Fellow
Demon	Lesson
Anger	Assist
Weapon	Cotton
Brutal	Bucket
Horror	Logic

APPENDIX C: MANIPULATION CHECK QUESTIONS FROM STUDY 2

1. How bothered were you that the reading passages were timed? ($M = 4.39$, $SD = 2.60$)
2. How bothered were you that the test was computer adaptive? ($M = 4.59$, $SD = 2.72$)
3. How bothered were you that your performance on the reading comprehension task is an indicator of intelligence and academic success? ($M = 4.89$, $SD = 2.88$)
4. How bothered were you that your performance on the reading comprehension task was videotaped? ($M = 3.82$, $SD = 2.65$)
5. How much did the task instructions (i.e., being videotaped, timed reading passages, computer adaptive testing, performance is an indicator of intelligence and academic success) influence your performance? ($M = 4.85$, $SD = 2.44$)
6. On similar exams (e.g., SATs, ACTs) in the PAST, how much have these type of instructions (i.e., being videotaped, timed reading passages, computer adaptive testing, performance is an indicator of intelligence and academic success) influenced your performance? ($M = 5.82$, $SD = 2.48$)

APPENDIX D: SAMPLE PASSAGE AND COMPREHENSION QUESTIONS

The ancient Egyptians believed strongly in life after death. They also believed that a person would need his body to exist in this afterlife. Therefore, they carefully preserved the body by treating it with spices and oils and wrapping it in linen cloth. The wrapped body was then placed in a tomb. A body that is treated in this way is called a mummy.

Egyptian kings and nobles wanted to be certain that their mummies would be kept in safe places forever. They had great tombs built for themselves and their families. Many kings were buried in secret tombs carved out of solid rock in a place near Thebes called the Valley of the Kings.

About eighty kings built towering pyramid-shaped stone tombs. These pyramids have become famous as one of the Seven Wonders of the Ancient World.

One of the most amazing things about the pyramids is that they were constructed without using wheels or heavy equipment to move or raise the rocks. Egypt did not learn about the wheel until long after the pyramids were built. Workmen used levers to get large blocks of stone on and off sledges and hauled them into place over long ramps built around the pyramids.

1. The term “mummy” was used to describe

- (A) Kings of ancient Egypt
- (B) Ancient Egyptian nobles
- (C) The place where Egyptian kings were buried
- (D) *The preserved body of a dead person*
- (E) One of the Seven Wonders of the Ancient World

2. The pyramids were built

- (A) *before the Egyptians developed a sophisticated technology*
- (B) after the Egyptians developed a sophisticated technology
- (C) to house the tombs of all ancient Egyptian kings and nobles
- (D) with the use of spices, oils and linen cloth
- (E) to keep mummies safe forever

3. Which of the following practices is most closely associated with ancient Egyptian belief in an afterlife?

- (A) placing the dead in tombs carved out of solid rock
- (B) building pyramids to house the bodies of dead kings
- (C) *preserving dead bodies with oils and spices*
- (D) creating the Valley of the Kings near Thebes
- (E) constructing tombs without the use of wheels or heavy equipment